



## The Transcriptional Landscape of the Production Organism *Pseudomonas putida*

D'Arrigo, Isotta

*Publication date:*  
2016

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
D'Arrigo, I. (2016). *The Transcriptional Landscape of the Production Organism Pseudomonas putida*. Novo Nordisk Foundation Center for Biosustainability.

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# **The Transcriptional Landscape of the Production Organism *Pseudomonas putida***

Ph.D. Thesis

**Isotta D'Arrigo**

The Novo Nordisk Foundation Center for Biosustainability  
The Technical University of Denmark

April 2016



**The Transcriptional Landscape of the Production Organism**  
*Pseudomonas putida*

**Ph.D. Thesis 2016 © Isotta D'Arrigo**

Novo Nordisk Center for Biosustainability

Technical University of Denmark

Kogle Alle 6, DK - 2970 Hørsholm

Denmark

Cover art: image adapted by Se Hyeuk Kim.

*“It’s not what you look at that matters, it’s what you see”*

(Henry David Thoreau)

*To my grandfather,*

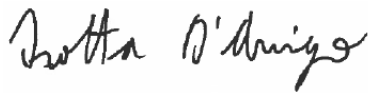
*I did not know why I am here*



## **Preface**

This thesis is written as a partial fulfillment of the requirements to obtain a Ph.D. degree at the Technical University of Denmark. The work presented here was performed between May 2013 and April 2016 at the Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark. The work was supervised by Katherine S. Long, Associate Professor from the Technical University of Denmark. Funding was provided by The Novo Nordisk Foundation and a Ph.D. grant from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme, [FP7-People-2012-ITN], under grant agreement no. 317058, "BACTORY".

The thesis was evaluated by Professor Juan Luis Ramos Martin from Abengoa Research, by Jeppe Vinther, Associate Professor at the University of Copenhagen, and by Alex Toftgaard Nielsen, Professor at Technical University of Denmark.



---

Isotta D'Arrigo

Hørsholm, April 2016

## Acknowledgements

First, I would like to thank my supervisor Katherine S. Long for her continuous support and motivation during my entire Ph.D. You have been an encouraging guide, who let me choose the direction but also drive me on the proper way.

Thanks to all the members (and ex-) of the RNA group, I cannot imagine how it would be without you! Special thanks to Holger, for inspiring me so many times and in so many occasions even without being at the center. You have been my first teacher, colleague and friend in the lab. I have learnt how to “bend the rules” thanks to you! Martin, for your generous patience and passion in making my things work. For all those days spent in front of a black screen with weird codes running...people watching me were thinking I was doing difficult programming work...I let them think of it, they did not know I was just trying to learn the “cd” command. A big thanks to Klara; even if you talk a lot, really a lot, now I can honestly say “it has been a big pleasure to work with you!” Together we worked in the RNA lab, together we crashed our “head” with bioinformatics in Český Krumlov, together we crashed my head in Cancun, and together we had also a lot of fun! Beside at work, I am so glad to have a friend like you...thanks a lot for everything! A huge thanks to Michelino, because you are not just a helpful and always happy co-worker but mainly a great boyfriend. Thank you for listening, for understanding me, for your continuous reinforcement. Thanks Miche, for always taking care of me... and for teaching me the proper use of “*by chance*”. Xiaochen, for all our discussions about work problems and scientific doubts, but at the end let's order pizza vah!

I would like to express my thanks to the Bactory people. Thanks to Søren and Mette, who built an amazing program and gave me the opportunity to take part in it. Thanks to all the Bactory Ph.D. students, it was a great pleasure starting with you that 1<sup>st</sup> of May three years ago, and since then we have always helped and strengthen each other. In particular many thanks to Mimi, for all your microbiology lessons, for your continuous support at work and in daily life, for all the laughs we had from Stockholm to China and for the pleasure of having you “endless” and singing Chandelier. Thanks to papiciulo Henrique, for your incredible help in every situation. You know *“Hey brother...if the sky comes falling down for you, There’s nothing in this world I wouldn’t do.”* Thanks to Giuggiolina for all our conversations, all our outbursts, and all our relaxing “cenette”. I will miss everything!

I would like to thank the Cfb collaborators João, Niko, Markus, Maja, thanks a lot for this final rush, I really appreciated all of your efforts. Thanks to the Cool Office and all the people who make it cool, it was awesome to spend these three years with you (Sofie, Patri, Dario)! I also would like to thank Carlotta who has always stood on my side. Thank you for our long talks, funny moments, and Elvira breaks! Big big thanks to Mr. Kim for all the help in the lab, with the figures, and with any kind of doubts. Always available to help, you are really awesome! Thanks to Laura, Virginia, Roberto, Alicia, Anna, Gigi, Nabin, Christian, Ida, Rosa, Math because all of you guys have helped me to enjoy these three years at Cfb. I would like to thank Allan and Anne from Novozymes, for hosting me in your department. It has been a wonderful experience with you!

I also would to thank the Italian group here in Copenhagen. Thanks to my brother Niccolò, who bravely took a decision and started a new adventure here. My historical friends Giulia and Isacco; I am closing



another circle and I will bring you in the next one. Thanks to Marchino for all the funny moments and Eugene. My flatmate Nicola; it has been great to know you, even if you support Djokovic.

I would like to thank my mum, dad, and grandmother Erichina, because even with the distance I have felt your strong support. You have always believed in me and push me to follow my dreams. Thanks a lot for that! Finally, I would like to thank all my friends in Italy and abroad that never give up on me. I cannot mention you all but you are amazing!

## Abstract

Bacterial cell factories represent a valid alternative to fossil fuel-based production. A promising bacterium that can be optimized as cell factory is *Pseudomonas putida*. However, its development in bioproduction applications poses some challenges including a clear understanding of the bacterial system biology.

This thesis has the aim of facilitating the development of *P. putida* KT2440 as a bacterial cell factory by investigating the transcriptome of the bacterium under different conditions (e.g. growth and stress). The main goals are the identification of differentially expressed genes, which provide information on bacterial adaptation to different environments, and the identification of non-coding RNAs, which regulate gene expression. This work focuses on several aspects of *P. putida* highlighting genomic features such as transcription start sites (TSSs), RNA regulatory elements such as riboswitches and small RNAs (sRNAs), metabolic pathways and transporter systems.

The results reported here significantly increase knowledge of the *P. putida* transcriptome, adaptation mechanisms, and reveal novel bacterial features that will aid the design and optimization of the bacterium as a cell factory.

## Dansk resumé

Bakterielle cellefabrikker udgør et reelt alternativ til produktion baseret på fossile brændstoffer. En lovende kandidatbakterie, der kan optimeres som cellefabrik, er *Pseudomonas putida*. Brugen af denne bakterie til bioproduktion har dog nogle udfordringer, hvilket inkluderer forståelsen af det bakterielle system.

Denne afhandling har som mål fremme udviklingen af *P. putida* KT2440 som en bakteriel celle fabrik ved at undersøge transkriptomet af bakterien under forskellige forhold (fx vækst og stress). De overordnende mål er identifikation af differentielt udtrykte gener, hvilket giver information om den bakterielle adaptation til forskellige miljøer, og identifikation af ikke-kodende RNA'er, der regulerer genekspression. Arbejdet fokuserer på adskillige aspekter af *P. putida* og fremhæver genomiske egenskaber som transkriptionelle startsteder (TSSs), regulatoriske RNA elementer som 'riboswitches' og små RNA'er (sRNAs), metaboliske netværk og transportsystemer.

Resultaterne, der her rapporteres, øger markant forståelsen af *P. putidas* transkriptom, tilpasningsmekanismer og beskriver bakterielle egenskaber, der vil hjælpe med designet og optimeringen af denne bakterielle cellefabrik.

## **Publications**

### *Included in this thesis*

**I. D'Arrigo**, K. Bojanovič, X. Yang, M. H. Rau, K. S. Long. (2016). Genome-wide mapping of transcription start sites yields novel insights into the primary transcriptome of *Pseudomonas putida*. *Environmental Microbiology*. Manuscript in press.

K. Bojanovič, **I. D'Arrigo**, K. S. Long. (2016). Global transcriptional responses to oxidative, osmotic, and imipenem stress conditions in *Pseudomonas putida*. Manuscript in preparation.

**I. D'Arrigo**, J. G. R. Cardoso, M. Rennig, N. Sonnenschein, M. J. Herrgård, K. S. Long. (2016). Investigation of *Pseudomonas putida* KT2440 transcriptome in different carbon sources reveals novelty in bacterial uptake systems. Manuscript in preparation.

### *Not included in this thesis*

H. Machado\*, A. M. Cavaleiro\*, **I. D'Arrigo**, K. Bojanovič, L. Gram, M. H. H. Nørholm. (2016). Exploring marine environments to unravel tolerance mechanisms to relevant compounds. Manuscript in preparation.

\*(These authors contributed equally)

## Content

Preface.....	i
Acknowledgements .....	ii
Abstract .....	v
Dansk resumé .....	vi
Publications .....	vii
Content .....	viii
Introduction and thesis outline .....	1
1. The soil bacterium <i>Pseudomonas putida</i> .....	3
1.1 The metabolic versatility of <i>P. putida</i> .....	4
1.1.1 <i>P. putida</i> grows on different carbon sources .....	4
1.1.2 Stress tolerance of <i>P. putida</i> .....	5
1.2 The potential of <i>P. putida</i> as a microbial cell factory .....	8
1.3 <i>P. putida</i> KT2440 as model organism.....	11
1.3.1 Genomic features of <i>P. putida</i> KT2440 .....	12
1.3.2 Metabolism of <i>P. putida</i> KT2440.....	13
2. Bacterial transcriptomics .....	19
2.1 Advances in transcriptomic technologies .....	20
2.2 Library preparation for RNA-seq .....	27
2.3 Exploration of the bacterial transcriptome .....	32
2.3.1 Non-coding RNA regulators control gene expression....	34
2.4 Transcriptomic analysis of <i>P. putida</i> KT2440 .....	37
3. Concluding remarks and perspectives .....	39
4. Bibliography .....	41
5. Research Articles.....	55

## **Introduction and thesis outline**

Currently, energy generation and production of bulk chemicals rely mostly on fossil fuels. However, fossil fuels are non-renewable resources and their use raises serious environmental concerns. One alternative is to secure a sustainable production with a limited environmental impact by utilizing microbial cells as “factories” for production of commercially valuable products.

The development of efficient bacterial cell factories requires modifications of their metabolic system to improve the productivity and tolerance to the environmental stresses encountered during the production process. With this perspective, an important step is the host selection. A bacterial cell pre-endowed with traits such as metabolic versatility and stress tolerance, becomes a valuable candidate for further optimizations. Moreover, understanding the bacterial characteristics, like optimal growth conditions, ability to use alternative energy sources, metabolic reactions and regulatory mechanisms, helps elucidating relevant features for the design of the bacterial cell factory. In this sense, the development of “omics” technologies enabled a comprehensive picture of bacterial systems by mapping information of genes (genomics), analyzing RNA molecules (transcriptomics), studying protein structure and function (proteomics), and revealing the network of chemical processes (metabolomics). At the same time, synthetic biology tools are continuously being developed to allow engineering of cell factories. The goal is to improve the host’s natural properties or introduce new ones (e.g. the production of a specific molecule) enabling the final design and optimization of the cell factory.

This Ph.D. thesis focused on understanding distinctive traits of *Pseudomonas putida*, one of the most prominent emerging bacterial cell factories. Using a transcriptomic approach, the set of RNA molecules expressed in *P. putida* under different growth conditions was investigated. The overall goal of this Ph.D. work was to facilitate the development of production strains by understanding *P. putida* adaptation to different growth and stress conditions, and by revealing RNA elements involved in gene expression regulation.

This thesis contains two main introductory sections. **Section 1** focuses on highlighting qualities of *P. putida* that make the bacterium interesting for biotechnology applications. Hence, the ability of the bacterium to adapt and survive in different environments (e.g. growth and stress conditions), and its capability to synthesize bioactive compounds, are disclosed. Examples of its strengths to deal with available growth resources and to endure stress factors are reported, as well as instances of its fruitful applications in synthesizing compounds of industrial relevance. At the end of Section 1, attention is narrowed to *P. putida* KT2440 (the strain used throughout this Ph.D. work); genomic features and metabolic pathways, relevant for this work, are described. **Section 2** introduces the general theory of transcriptomics, and gives an overview of recent technological advances that enabled the development of this field of research. The section also points out the spectra of information that is possible to obtain using transcriptomic approaches (e.g. genes expression, gene structure and RNA regulatory elements). Finally, Section 2 re-connects with Section 1 via a brief analysis of results from transcriptomic studies of *P. putida* KT2440.

Concluding remarks and future perspectives are then presented, which lead to the presentation of the research articles generated with the results obtained in this Ph.D. study.

## **1. The soil bacterium *Pseudomonas putida***

The genus *Pseudomonas* belongs to the  $\gamma$ -Proteobacteria, and includes a large number of Gram-negative bacterial species (>200 in the Approved Lists of Bacterial Names [1]). Bacteria belonging to this group share the ability to adapt to different environmental niches, endure stress conditions, and synthesize several bioactive compounds. *P. putida* and *P. aeruginosa* are the most studied models in the genus. Both species are present in many environments (e.g. soil, water), and several *P. putida* strains also colonize the rhizosphere by establishing beneficial relationships with plants [2]–[4]. Instead, *P. aeruginosa* is often found in humans as an opportunistic pathogen that infects immunocompromised patients (e.g. in lung infections of cystic fibrosis patients) [5]–[7].

The ability of *P. putida* to synthesize compounds [8], together with the lack of virulence factors [8], [9], make the bacterium appealing for the development of efficient production strains. Moreover, the capability to degrade pollutants and toxic substrates make it a suitable candidate as biocatalyst in bioremediation [10]. For these reasons, many studies have focused on *P. putida* in the last decades and many aspects of its biology have been revealed, while others remain to be elucidated.

In this section, the metabolic features that make *P. putida* a valid candidate for biotechnological applications are reviewed. Examples of its ability to adapt to different environments (e.g. growth and stress conditions), and its possible applications in production are reported. Particular focus is given to *P. putida* KT2440, the most prominent and well-studied *P. putida* strain, which has been used in this Ph.D. work.



## **1.1 The metabolic versatility of *P. putida***

*P. putida* is a ubiquitous rod-shaped, flagellated Gram-negative bacterium. It can live in different environmental niches due to its metabolic versatility, low nutritional requirements, and high tolerance to toxic compounds [11]–[13]. The bacterium is usually found in soil or water ecosystems, and in the plant rhizosphere, where it imparts a positive effect on plants by stimulating growth (e.g. by synthesizing the plant growth factor auxin), and suppressing pathogens (e.g. by secretion of bioactive factors that attack the pathogens) [14], [15]. Its colonization niches vary greatly in their composition, containing mineral and organic compounds, gases, water, liquids and are often characterized by continuous physio-chemical changes. Therefore, *P. putida* has developed mechanisms to exploit the available nutrients and to endure stress variations. Indeed, *P. putida* is able to monitor a variety of environmental signals, and subsequently adapt its physiological status via complex regulatory networks that control cellular metabolism [12].

### **1.1.1 *P. putida* grows on different carbon sources**

*P. putida*, as the other *Pseudomonas* spp., is able to assimilate a wide range of compounds, and differently from *Escherichia coli* and *Bacillus subtilis*, prefers to use organic acids (e.g. succinate) and amino acids abundant in its environment as carbon sources rather than glucose. In the presence of both succinate and glucose, *P. putida* reduces the assimilation of glucose and promotes a metabolism to use its preferred compound more rapidly [16], [17]. The regulatory mechanism that allows cells to select and use the preferred carbon sources in a mixture is called carbon catabolite repression [18]. However, *P. putida* is still able

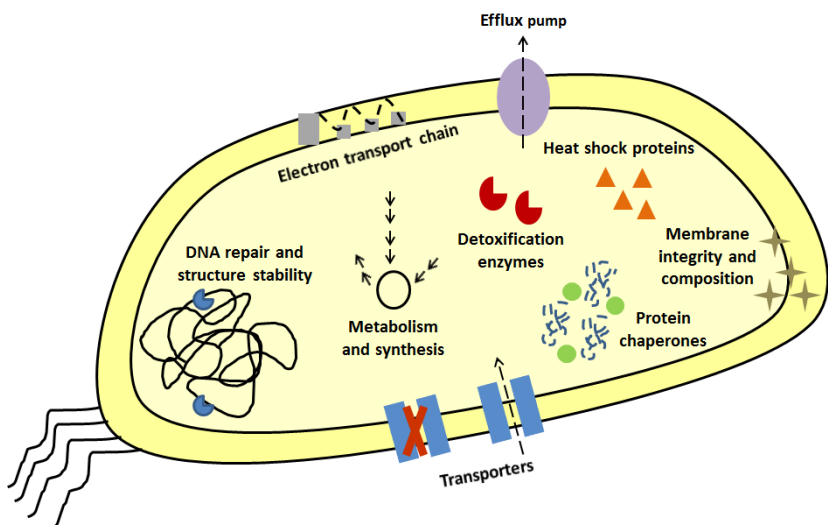
to growth well on sugar substrates such as glucose, gluconate, glycerol, fructose and mannitol [16], [17], [19].

*P. putida* can use amino acids as carbon sources, and the assimilation of them is also regulated by the carbon catabolite repression mechanism. Indeed, when growing on a mixture of them, the bacterium selects preferred amino acids first. *P. putida* prefers to use proline, alanine, glutamate, glutamine and histidine and represses the assimilation of valine, isoleucine, leucine, tyrosine, phenylalanine, threonine, glycine and serine [20].

Moreover, *P. putida* has the capability to use aromatic compounds, mostly deriving from decomposition of plant material, as energy sources. Some of such compounds used by most *P. putida* strains are benzoate, *p*-hydroxybenzoate, phenylacetate, tyrosine, phenylalanine, benzylamine, nicotinate, ferulate, coumaryl alcohols and quinate [2], [21]–[23]. Carbon catabolite repression also regulates the catabolism of aromatic compounds; for instance, in *P. putida* the genes involved in benzoate and *p*-hydroxybenzoate assimilation are repressed when the bacterium grows in a rich media supplemented with benzoate and *p*-hydroxybenzoate [24], [25].

### **1.1.2 Stress tolerance of *P. putida***

In its natural habitats, *P. putida* copes with stress situations and extreme conditions such as temperature variations, extreme pH values, or the presence of toxins, inhibiting solvents or pollutants [13]. Therefore, *P. putida* needs a metabolic apparatus to adapt to changes for prolonged periods of time (Figure 1).



**Figure 1. *P. putida* stress tolerance.** *P. putida* endures stress conditions by stabilizing DNA, RNA and protein structures, altering the electron transport chain, adjusting metabolism and synthesis pathways, activating detoxification enzymes, and modifying the membrane (e.g. transporters, efflux pumps and membrane composition).

In the past years, studies have been done on *P. putida* with the aim of understanding the response and adaptation of the bacterium to different stress conditions. Studies of *P. putida* growing at low temperatures (4-10°C) have showed that the bacterium modifies the expression of genes involved in energy metabolism and in transport. Moreover the bacterium responds to increased protein aggregation and stability of DNA/RNA structures, and to the reduction of membrane fluidity by modifying lipid composition [26], [27]. At low pH, *P. putida* responds to the stress by retention of pH homeostasis, altering cell division, the composition of the outer membrane, the electron transport chain, and by regulating the

heat-shock response [26]. Oxidative stress is also often encountered by soil bacteria, as a consequence of interaction with environmental agents that induce reactive oxygen species (ROS). To sustain oxidative stress, *P. putida* regulates a network of stress-regulator and induces the expression of stress-sensing proteins and detoxification enzymes [28]–[36]. Soil microorganisms are also often affected by water stress, which refers to desiccation when promoted by osmotic agents, or instability of hydrogen bonds in water solution when promoted by chaotropic solutes (e.g. pollutants). Osmotic agents influence cell turgor and consequently *P. putida* responds by regulating transport systems, inducing membrane integrity and the synthesis of cytosol solutes [37]–[39]. Chaotropic solutes reduce water activity and perturbs macromolecular interactions, leading the cell to upregulate proteins involved in protein-structure stabilization, and alter lipid metabolism and membrane composition [40].

Finally, organic solvents (e.g. aromatic and aliphatic toxic compounds) can accumulate in the soil and affect *P. putida* cells. *P. putida* is equipped with metabolic pathways, encoded by chromosomal genes, for the degradation of toxic pollutants [23]. Moreover, some isolates harbor plasmids containing the genetic information required for the breakdown of toluene (3-methyl-benzoate) [41], naphthalene [42], 4-chloronitrobenzene [43], 2,4-xyleneol [44], or phenol [45]. The toxicity of aromatic substrates is mainly due to their accumulation in the membrane, which alters its stability and proper function. Moreover, aromatic compounds cause oxidative stress. *P. putida* responds by activating efflux pumps to extrude the solvent, inducing chaperones to maintain protein stability and ROS-scavengers against oxidative stress, and altering the composition of the membrane [46].

The capability to tolerate and modify toxic organic substrates allows *P. putida* to have a significant impact in environmental applications such as bioremediation. Owing already a natural apparatus for the detoxification of toxic compounds, *P. putida* could be engineered for the developed of strains with superior tolerance characteristics.

## **1.2 The potential of *P. putida* as a microbial cell factory**

In recent years, the potential of *P. putida* as a cell factory has started to be exploited and the bacterium has showed excellent production properties for a wide-range of natural and industrial compounds. *P. putida* has all the necessary features to be one of the major protagonists in the shift from oil-derived manufacturing to biosustainable production strategies. Indeed, the bacterium is endowed with production skills of secondary metabolites of biotechnological interest, and capability of *de novo* synthesis and biotransformation of high value chemicals and pharmaceuticals [8], [10], [47].

Of particular interest is the production of the group of secondary metabolites called polyhydroxyalkanoates (PHA). Depending on the growth condition, *P. putida* can deal with an excess of carbon, compared with other essential elements (e.g. nitrogen or phosphorus), and under these circumstances, accumulates granules of PHA as energy storage [48]. PHAs are linear polyesters with excellent biodegradability and biocompatibility properties that are interesting for applications in tissue engineering or in the production of bioplastics [49]. Based on the length of the monomer which constitutes the chain, PHAs can be classified in three groups: short-chain length (scl) with 3 to 5 carbon atoms, medium-chain length (mcl) with 6 to 14 carbon atoms and long-chain length (lcl)

with more than 14 carbon atoms. The monomeric composition affects PHA properties including elasticity, crystallinity, and rigidity, and mcls have improved elastomeric properties compared with scls [50]. Briefly, PHA biosynthesis occurs in two steps: (i) generation of the hydroxyacyl-CoAs (HA-CoAs) from various metabolites as precursors by  $\beta$ -oxidation, or from non-fatty acid precursors by fatty acids biosynthesis; (ii) polymerization of HA-CoAs in PHAs by Pha synthases [51], [52]. Many microorganisms (e.g. *E. coli* [53]) accumulate scl-PHAs, while *P. putida* can produce mcl-PHAs [54]. Moreover, *P. putida* has been optimized as a platform to produce mcl-PHAs from C4 to C14 fatty acids, by generating HA-CoAs with the same carbon number of the precursor fatty acid (C6-C14) [55], [56].

*P. putida* also shows great potential as a biocatalyst, in the bioconversion of aliphatic, aromatic, and heterocyclic compounds to high value products [8], [47]. As an example of biocatalysis, *P. putida* can metabolize benzoate and produce the intermediates *cis-cis* muconate, a precursor of adipic acid used for nylon production [57]. The bacterium can also convert ferulate to vanillate, a flavor agent used in food, beverages, and in the production of pharmaceuticals [58], [59], and convert styrene oil into the biodegradable plastic PHA [60].

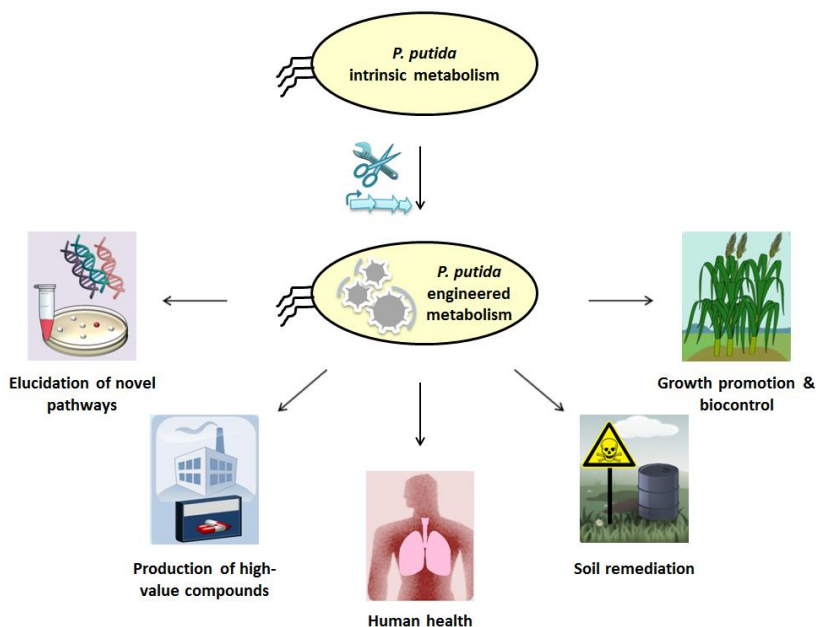
Finally, *P. putida* has shown great capability in *de novo* synthesis of natural compounds by expression of heterologous genes [47]. An interesting example of biosynthesis is the organic compound *p*-coumarate, which has been found in barley, wine and vinegar, and many edible plants such as peanuts, navy beans, tomatoes, carrots, and garlic [61]. In addition, *p*-coumarate is the precursor of phenylpropanoids [61], and also important in medical applications such as degradable plastic, orthopedic matrix, and drug delivery systems [62]. Purification of *p*-coumarate from plants is time-consuming [63] and chemical production

is expensive. Therefore new solutions to obtain the compound are needed. Via metabolic engineering approaches, *P. putida* has been enhanced to produce *p*-coumarate from glucose via the metabolite L-tyrosine. First, the bacterium has been engineered with the introduction of the phenylalanine/tyrosine ammonia lyase gene, followed by increasing the availability of the substrate tyrosine, preventing *p*-coumarate degradation and avoiding accumulation of its by-product cinnamate [64]. *P. putida* is also promising as a host organism for the expression of biosynthetic genes for synthesis of natural products from myxobacteria. Myxobacteria are root-living microorganisms that produce several secondary metabolites with various biological activities [65]. The production of myxochromide S has been established in *P. putida* by the expression of a *Stigmatella aurantiaca* biosynthetic gene cluster [66].

In conclusion, the simplicity of growth in a wide range of conditions, the metabolic skills that enable stress tolerance, and the capacity to produce industrially relevant compounds, make *P. putida* a desired organism for laboratory works with great potential in biotechnology applications. These natural hallmarks together with the engineering tools for genetic manipulation and gene expression now available for *P. putida* [9], allow this bacterium to be an easy and accessible model for future development in bioremediation and production.

Furthermore, *P. putida* is also considered for its applicability as a delivery system for bioactive compounds (e.g. outer membrane vesicles could be used as vaccine carrier) [67], and for plant growth promotion and plant protection (Figure 2) [68]. At the moment, the limiting step to an extensive application of *P. putida* is a lack of knowledge of bacterial behavior under industrial and environmental conditions, and the still limited toolbox for strain engineering and manipulation. Therefore, more

understanding of bacterial physiology, like that obtained during this Ph.D. work, and improved engineering tools will help to implement *P. putida* for biotechnology applications.



**Figure 2. *P. putida* applications and future perspectives.** New engineering and expression tools, together with the increasing understanding of bacterial metabolism, will generate designed *P. putida* cells to use in several biotechnological applications. (Figure adapted from ref. [47]).

### 1.3 *P. putida* KT2440 as model organism

In this Ph.D. work the *P. putida* KT2440 strain has been used, as it is one of the best characterized and well-studied *P. putida* strains [69]. Its



genome was sequenced in 2002 [11] and recently re-annotated [70]. *P. putida* KT2440 is the plasmid-free derivative of the toluene-degrading *P. putida* (*arvilla*) mt-2 [71], [41]. The KT2440 strain is the first Gram-negative soil bacterium to be certified as a biosafe strain by the Recombinant DNA Advisory Committee (Federal Register, 1982)<sup>1</sup>.

### 1.3.1 Genomic features of *P. putida* KT2440

The KT2440 genome is a single circular chromosome of 6,181,873 bp in length with a GC content of 61.5%. Genome analysis has annotated 5592 open reading frames, of which more than 80 encode oxidoreductases and more than 40 encode oxygenases that are important in chemical transformations [11], [67], [70]. The genome contains around 30 porin-encoding genes, including *oprB-1* for glucose entry, and 23 members of the *benF/phaK/oprD* family for the uptake of aromatic compounds. *P. putida* KT2440 has approximately 370 cytoplasmic membrane transport systems, and the majority corresponds to ATP-binding cassette (ABC) transporters. In connection with its ability to colonize plant roots and use amino acids as energy sources, a great proportion of transporters are associated with amino acid uptake. The genome also encodes transporters for organic acids (dicarboxylic, tricarboxylic, and aromatic acids), and sugar uptake (e.g. glucose, gluconate, glycerol, ribose, fructose) [11], [70].

The genome encodes several osmoprotectant uptake systems, including ABC transporters for glycine and proline betaine, a proline betaine major facilitator superfamily (MFS) transporter, and six members of the choline/carnitine/betaine transporter family. *P. putida* KT2440 also

---

<sup>1</sup> Federal Register (1982) Appendix E, Certified host-vector systems. 47: 17197.

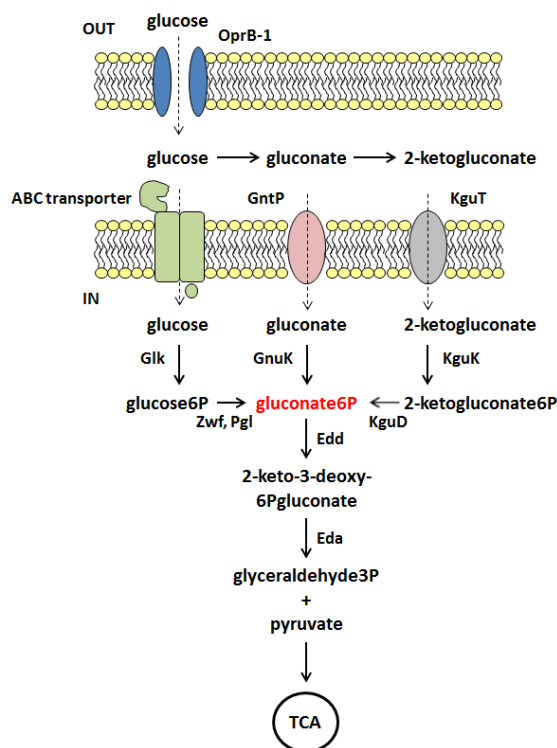
contains a large number of efflux systems families to extrude metals, organic solvents, and antibiotics [11], [13]. Among the different efflux pumps, the genome encodes the RND (Resistance-Nodulation-Division) efflux pumps TtgABC, MexCD-OprJ and MexEF-OprN [72]. TtgABC is the main responsible for antibiotics resistance (e.g.  $\beta$ -lactams, chloramphenicol and quinolones) in KT2440 [73], [74]. Moreover, TtgABC is also implicated in toluene [73] and in phenol resistances [75]. MexCD-OprJ and MexEF-OprN are also involved in antibiotics resistance [73], with the latter also implied in formaldehyde detoxification [76]. The KT2440 genome also encodes an ABC transporter Ttg2ABC, which may be involved in toluene [77] and phenol stress [75].

### **1.3.2 Metabolism of *P. putida* KT2440**

As considered before, *P. putida* KT2440 preferably uses organic acids, such as intermediates of the tricarboxylic acid (TCA) cycle, and amino acids as carbon sources, but it is also able to use a wider range of substrates, including sugars and aromatic compounds. Considering glucose as representative compound of the hexose class and also the most widely used carbon source in laboratory settings, the glucose catabolism in KT2440 is now described.

Differently from *E. coli*, *P. putida* KT2440 has an incomplete Embden–Meyerhof–Parnas (EMP) pathway for glycolysis as it lacks the gene for 6-phosphofructokinase (Pfk), which phosphorylates fructose-6-phosphate to fructose-1,6-bisphosphate [11], [78]–[81]. However, KT2440 can metabolize C3 and C6 sugars via the Entner–Doudoroff (ED) pathway (Figure 3) [13], [28], [78], [82], [83]. In the first step of hexose catabolism, glucose is transported in the periplasmic space

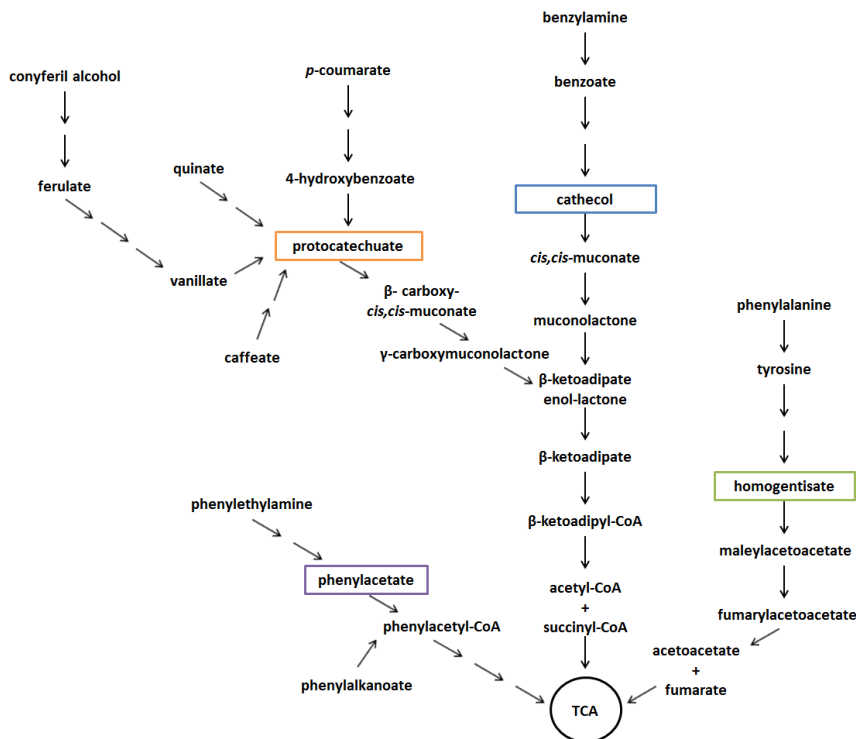
through the outer membrane porin OprB-1 [78], [84], [85]. Then, glucose can be directly transported by an ABC transporter system into the cytoplasmic space, where it is phosphorylated to glucose-6-phosphate by the kinase Glk. Alternatively, glucose can diffuse into the periplasm and be oxidated to gluconate and then to 2-ketogluconate by periplasmic dehydrogenases. The intermediates gluconate and 2-ketogluconate are then transported into the cell by specific transporter systems, the gluconate transporter GntP and the 2-ketogluconate transporter KguT, and subsequently phosphorylated by GnuK and KguK. The three upstream routes converge at the level of gluconate-6-phosphate, which enters the ED pathway. Finally, the key enzymes of the ED pathways, Edd and Eda, convert gluconate-6-phosphate into glyceraldehyde-3-phosphate and pyruvate. Pyruvate is then converted into oxaloacetate or acetyl-CoA, intermediates of the TCA cycle [78], [86].



**Figure 3. Glucose metabolism in *P. putida* KT2440.** Schematic representation of the glucose metabolism in *P. putida* KT2440. The three upstream routes (glucose, gluconate and 2-ketogluconate) converge at the level of gluconate-6-phosphate (red), which is then metabolized by the ED pathway.

As previously mentioned, *P. putida* KT2440 is able to grow in the presence of aromatic compounds, derived from plant materials, as sole carbon sources. Indeed, the bacterium can grow in minimal medium containing benzoate, *p*-hydroxybenzoate, benzylamine, phenylacetate, phenylalanine, tyrosine, phenylethylamine, phenylhexanoate, phenylheptanoate, phenyloctanoate, coniferyl alcohol, *p*-coumarate,

ferulate, caffeate, vanillate, nicotinate and quinate [23]. Four different central pathways, i.e. the catechol (*cat*), the protocatechuate (*pca*), the phenylacetate (*pha*) and the homogentisate (*hmg*), for the catabolism of aromatic compounds, have been predicted in the genome (Figure 4). The catechol and the protocatechuate pathways represent the two branches of the  $\beta$ -ketoadipate central pathway, and a common subset of enzymes (*pcaDIJF*) is involved in the final conversion to succinyl-CoA and acetyl-CoA [23].



**Figure 4. The four catabolic pathways for aromatic compounds.** The four central aromatic intermediates are shown in colored boxes; protocatchuate in red, catechol in blue, homogentisate in green and phenylacetate in purple. (Figure adapted from ref. [23]).

The predicted aromatic pathways and their role in catabolism of aromatic compounds, have been supported by proteomic analysis, which has showed the induction of the major aromatic enzymes under growth on benzoate, *p*-hydroxybenzoate, vanillin, phenylethylamine and phenylalanine as sole carbon sources [87]. Furthermore, several *in-silico* genome-scale metabolic analyses of *P. putida* KT2440 growing on

several carbon sources, including aromatic compounds, have been constructed, and have confirmed the ability of the bacterium to grow on a wide-range of substrates, and to degrade aromatic molecules [79], [80], [88], [70], [89]. However, even if many works have been addressed to understanding metabolic pathways and regulatory mechanisms of *P. putida* KT2440, many biological processes remain to be elucidated.

In summary, this section has presented the hallmarks of *P. putida* (e.g. versatility to survive in different environments, innate and acquired ability to synthesize compounds), which make the bacterium attractive for biotechnology applications. In particular, genomic characteristics and metabolic abilities, which underline these hallmarks, have been further described for the well-studied strain KT2440. Nevertheless, the section accentuates the necessity to obtain a deeper knowledge of *P. putida* KT2440 in order to develop the bacterium as cell factory. In this perspective, a key step is to understand the bacterial transcriptome, a topic that has been the focus of this Ph.D. work, and that is treated in the next section.

## 2. Bacterial transcriptomics

The central dogma of biology, proposed by Crick in 1958 [90], explains the flow of genetic information in biological systems: *DNA makes RNA and RNA makes protein*. In the past few decades, however, advances in the understanding of the genome, transcriptome, proteome, and metabolome have contributed to extend the central dogma and to shift its linearity to a more complex view. The revisited central dogma de-emphasizes the unidirectional flow of information from DNA to RNA and proteins, underlines the emerging properties for the single factors and the networks between them, and introduces metabolism as active player [91], [92]. In the new vision, RNA has also acquired more relevance. Evidence indicates that RNA is not only a messenger between DNA and proteins, but also exhibits structural diversity and biological activity. RNA can store genetic information, catalyze reactions and regulate gene expression and protein activity [93], [94], [95].

Transcriptomics is the study of the complete set of RNA transcripts (transcriptome) that are present in a cell at a specific moment in a certain growth condition. Transcriptomic studies focus on the identification of genes that are differentially expressed depending on the physiological status of the cell, and on the identification of RNA-based regulation mechanisms. In recent years, the development of high-throughput sequencing technologies has allowed a deep understanding of the bacterial transcriptome and revealed an unexpected level of complexity and diversity [96], [97].

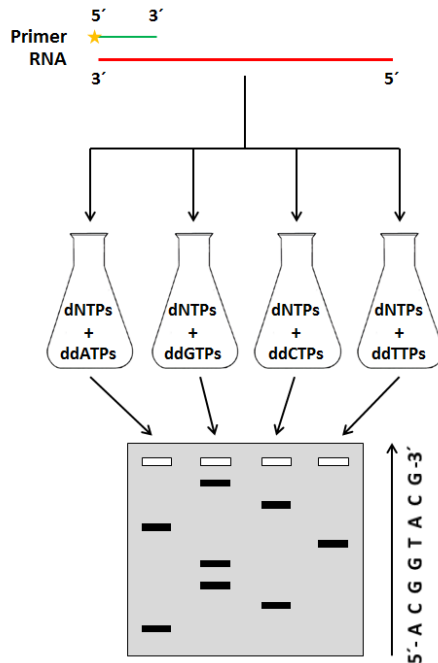
In this section, the major advances in transcriptomic technologies and new insights into genomic elements and regulatory RNA in bacteria, gained by transcriptome analyses, are reviewed. Finally, the section concludes with examples of transcriptome studies on *P. putida* KT2440.



## **2.1 Advances in transcriptomic technologies**

The study of the transcriptome is a qualitative and quantitative evaluation of the population of RNA transcripts expressed in a cell, which varies with external environmental conditions. It includes RNA transcripts, reflecting the genes actively expressed and non-coding RNA elements involved in many cellular processes (e.g. regulatory mechanisms). The study of the transcriptome also refers to expression profiling, in order to examine the expression level of RNAs in a cell.

The first sequencing of RNAs was performed by complementary DNA (cDNA) synthesis of RNA transcripts and Sanger sequencing [98]. In the Sanger method (Figure 5), invented in 1977, the target RNA is annealed to an oligonucleotide DNA primer radioactively labeled at the 5' end. Reverse transcriptase then extends the primer generating a cDNA. The supplied mixture of deoxynucleotide triphosphates (dNTPs) includes a chain-terminating dideoxynucleotide triphosphate (ddNTP). When the ddNTP is incorporated, the reverse transcriptase stops and the chain is terminated. The reaction is performed four times, using a different ddNTP for each reaction. The generated fragments are then separated on a gel matrix and the sequence information can be obtained from the migration order.

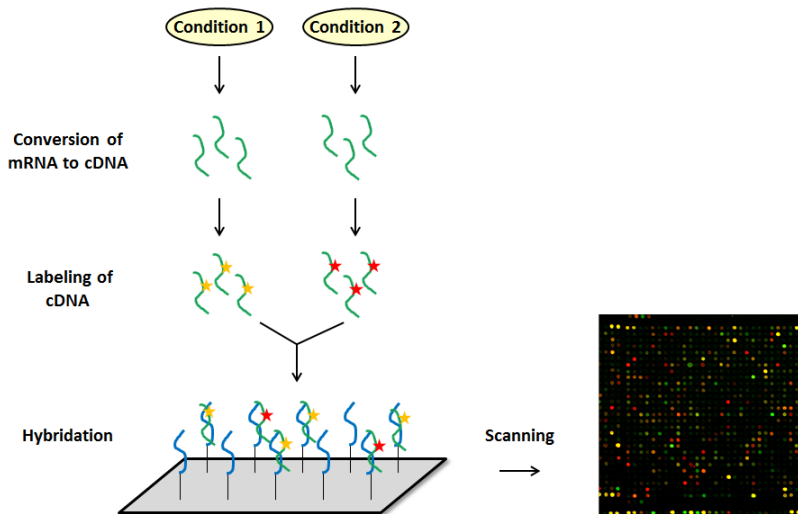


**Figure 5. Sanger sequencing.** A 5' radiolabeled primer anneals to the template RNA in four different reaction mixtures. Each reaction contains the reverse transcriptase, the four dNTPs, and one of the four possible ddNTP. The reverse transcriptase extends the primer and when the ddNTP is attached, the growing chain is terminated. The fragments generated in each reaction are separated on a gel matrix and the synthesized sequence is read thanks to the migration order. Colours stand for: oligonucleotide DNA primer (green line); RNA (red line). (Figure adapted from ref. [99])

One of the major problems of the Sanger sequencing was the premature termination of the cDNA molecules, due to the presence of secondary structures causing reverse transcriptase to pause. The original

Sanger method has been automated and improved by substituting the radioactivity labeling of the newly synthesized strand with fluorescent dye terminators. By using fluorescent dye terminators, the premature terminated cDNA molecules are not detected, and also the sequencing is possible in a single reaction [100]. However, this sequencing approach still has low throughput, is expensive and not quantitative. Therefore additional efforts have been made to develop new methods [101], [102].

In the late 1990s, the hybridization-based microarray technology was developed. Microarrays enable the quantification of expression of all annotated genes in a genome, and are also suitable to compare gene expression under different conditions (Figure 6). Annotated genome sequences are used to construct an oligo-probe microarray representing part of or the entire genome. The RNA sample is converted into a cDNA library, which is then fluorescence labeled and used for hybridization on the probes immobilized on the microarray. The hybridization between probes and target DNA is detected and quantified by fluorescence emission [103].



**Figure 6. Microarray technology.** Comparative analysis of gene expression from cells in different conditions is possible by labeling the cDNA libraries with different fluorescent dyes. Microarray spots with more labeled cDNA hybridized have more fluorescence intensity emission. The ratio of fluorescent intensities for a spot is correlated with differences in gene expression. Colours stand for: microarray probes (blue line); DNA (green line).

A subtype of traditional microarray, the so called tiling array, was then developed. Tiling arrays are high-density microarrays, and as the original microarray technology, function by hybridizing labeled DNA on the chips probes. However, tiling arrays differ from microarrays in the nature of the probes. Instead of using probes representing annotated genes, tiling arrays use short fragment probes covering the entire genome or contiguous regions of the genome, and are spotted in high density on the array. By covering the entire genome, tiling arrays have allowed for detection of novel genes and RNA transcripts in intergenic regions.

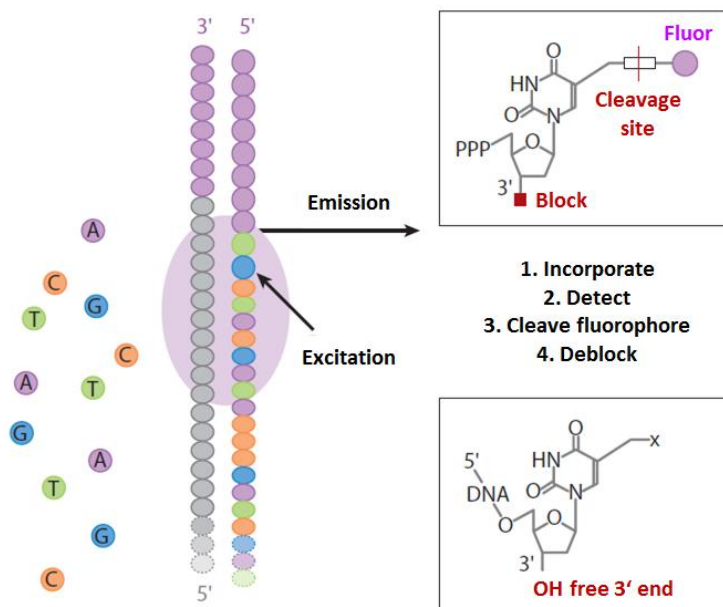
Moreover, detection of genomic features such as polymorphisms is possible by using overlapping probes for the same genomic region [104], [105]. Recently, tiling arrays have been applied to study several bacterial transcriptomes like *E. coli* [106], *Caulobacter crescentus* [107], *Listeria monocytogenes* [108], and *B. subtilis* [109], and have contributed to reveal new aspects of the transcriptome such as non-coding RNAs and antisense RNAs [108], [109].

However, microarrays have some limitations including: (i) a limited range of quantitative detection due to background noise, saturation and spot density, (ii) not being suitable for multiple species (or strains) analysis as mismatches on probes can affect the hybridization, (iii) requiring normalization methods to compare different assays, and (iv) the impossibility to distinguish between *de novo* synthesized transcripts and modified transcripts [110].

The revolution in sequencing occurred in 2008, when the high-throughput next generation sequencing (NGS) technologies were introduced. NGS technologies enable RNA samples analysis through their conversion in cDNA libraries and sequencing on a massive scale (RNA-seq). In RNA-seq, the cDNA libraries are covalently linked with adapters, universal sequences specific to each platform, and are attached, amplified and sequenced on a solid surface [99]. NGS is the term used to describe a number of different modern sequencing technologies including: Roche 454, Illumina (Solexa), SOLiD, and Ion Torrent Proton/PGM. Without going into too much detail, the Illumina approach is briefly described as it has been used during this Ph.D. work.

Illumina technology makes use of a flow cell, which contains covalently attached oligonucleotide sequences complementary to the library adapters. Library fragments are hybridized to the oligonucleotides on the flow cell surface, and amplified *in situ* by a bridge amplification

step to produce fragment clusters for sequencing. At the end of the amplification, only one strand remains attached to the flow cell, while the other is washed away. Then, the polymerase and the four fluorescent-labeled dNTPs are introduced on the flow cell to amplify the remained attached strand. The polymerase incorporates only one nucleotide per cycle as dNTPs own a blocking group at the 3'-OH position of the sugar. After the nucleotide is incorporated, the fluorescence is detected and the blocking group is removed to restore the free 3'-OH and allows the incorporation of the next dNTP. Therefore, Illumina works with subsequent cycle of amplification, sequencing, and analysis (“sequencing-by-synthesis” method) [99] (Figure 7).



**Figure 7. Illumina sequencing.** Representation of the main steps of the Illumina sequencing: incorporation of dNTP, detection of fluorescent, cleavage of fluorophore, deblock of 3'-OH position. (Figure adapted from ref. [99]).

The advantages of RNA-seq are the single-base resolution, the possibility to detect transcripts that do not correspond to existing genomic sequence, the low background signal, a large range of expression (no detection limit for quantification), the low cost compared with tiling arrays, and finally RNA-seq is not strain-specific [111]. However, also RNA-seq methods have some biases. The protocol for library preparation (see below) can be tricky, as several steps (e.g. mRNA enrichment, fragmentation, reverse transcription, adapter ligations, PCR) are required to finally generate the cDNA library; the sequencing is therefore subject to the biases inherent to these procedures.

Moreover, the read lengths obtained with the sequencing are limited by the signal to noise ratio, therefore sequencing technologies require enforced cDNA lengths as input for the sequencing analysis. However, several technologies have been able to increase the read length over time. Finally, RNA-seq requires the support of bioinformatic tools to map the reads to the genome, normalize the data, and analyze differential expression [99], [112].

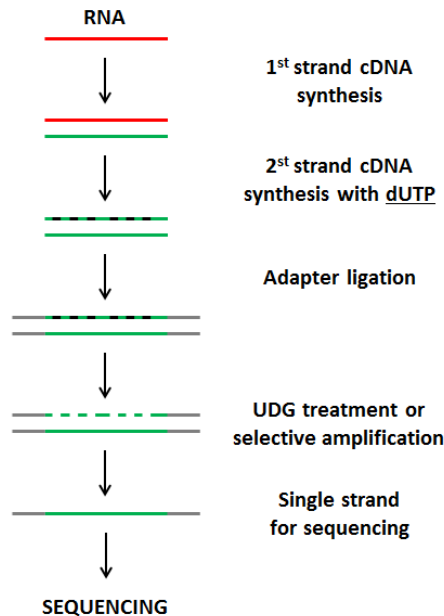
## **2.2 Library preparation for RNA-seq**

The bacterial transcriptome comprises several RNA species including: mRNA codifying proteins, rRNAs, tRNAs, small regulatory RNAs (sRNAs), and regulatory elements such as riboswitches. Expression levels of mRNA, operon composition, transcriptional start site, and non-coding regulatory RNA elements, are all important features of the transcriptome. These features can be specifically investigated in an RNA-seq experiment by designing a suitable cDNA library that leads to the generation of optimal sequencing data for the desired transcriptome elements. The standard library preparation involves the selection of an appropriate RNA fraction (e.g. total RNA or enriched mRNA), and its conversion into a cDNA library. A fragmentation step is performed before or after the cDNA preparation, which is required to respect the cDNA lengths imposed by sequencing technologies (see above). Adapter sequences are then linked to the cDNA fragments, and finally the library is PCR amplified and sequenced. The standard protocol for library preparation can be modified in several ways. Here, two recent improvements that have been used in this Ph.D., are described: the



strand-specific and the differential RNA-seq (dRNA-seq) approaches [113].

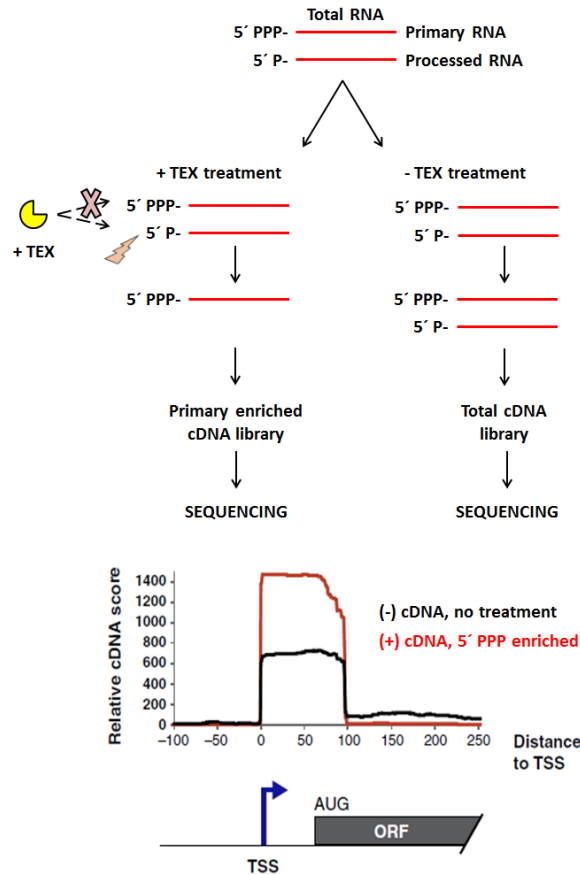
The strand-specific approach allows the determination of the specific DNA strand from which the RNA molecules originated, and it is suitable for identifying antisense transcripts, which are likely to be involved in regulation of gene expression. Among the methods to maintain the original strand information, marking the second strand with dUTP is often the protocol of choice [114], [115] (Figure 8). In this process, the synthesis of the second strand of cDNA is carried out in the presence of dUTP instead of dTTP. The result is a double stranded cDNA containing a first strand with deoxythymidine and a second strand with deoxyuridine residues. After the adapter ligation step, the second strand is selectively removed by degradation with uracil DNA glycosylase (UDG), or by using a DNA polymerase that cannot incorporate dUTP during amplification. Only the first strand is sequenced, and the strand information is maintained [116]–[120].



**Figure 8. The strand-specific dUTP method.** During the synthesis of the second strand of cDNA, dUTPs are used instead of dTTPs. This allows the selective removal of the UTP-containing cDNA strand by UDG treatment or PCR amplification. Colours stand for: RNA (red line); DNA (green line); adapter (grey line). (Figure adapted from ref. [115]).

The dRNA-seq strategy is a recently developed RNA-seq protocol to globally determine transcription start sites (TSS) of transcripts (Figure 9) [121]. The bacterial RNA pool consists of primary and processed RNAs, including the abundant rRNAs and tRNAs. Primary RNAs are characterized by the presence of a 5' triphosphate end (5' PPP), while processed RNAs contain a 5' monophosphate end (5' P). In the dRNA-seq protocol the total RNA sample is split in two aliquots; one aliquot is treated with 5' P-dependent terminator exonuclease enzyme TEX that

specifically degrades the 5' monophosphate transcripts (processed RNAs) but not the 5' triphosphate transcripts (primary RNAs). The second aliquot is not treated. The two RNA samples are then converted into cDNA resulting in a primary RNA enriched library (+TEX) and a total RNA library (-TEX). After sequencing, the primary enriched library shows an increased read coverage at the 5' end of transcripts compared with the total library. Therefore, a complete sequencing comparison of the two libraries allows for genome-wide mapping of TSSs. Moreover, dRNA-seq provides information about operon structures, promoter regions, 5' untranslated region (UTR) lengths, novel open reading frames (ORFs), antisense transcription and enables sRNA identification [122].



**Figure 9. The dRNA-seq approach.** Schematic representation of the construction of the two libraries: primary enriched cDNA library (+TEX, on the left) and the total cDNA library (-TEX, on the right). The comparison of read coverage at the 5' end of transcripts between the two libraries allows identification of TSSs. Colour stand for: RNA (red line). (The read profile is adapted from ref. [122]).

The first primary transcriptomic study using the dRNA-seq approach was performed on *Helicobacter pylori* [121]. A total of 1,907 TSSs were identified, including hundreds of transcriptional start sites within operons, and opposite to annotated genes. Moreover, the approach led to the discovery of ~ 60 sRNAs [121]. Since its introduction in 2010, dRNA-seq has been used to characterize many different bacterial transcriptomes [123]–[128]. During this Ph.D., we applied the dRNA-seq approach to study the primary transcriptome of *P. putida* KT2440 (Research Article 1) and we were able to identify 7,937 TSSs, where the majority were located on the opposite strand or internal to annotated genes. We also used the information of mapped TSS to investigate 5' UTRs and identified non-coding RNA elements.

### **2.3 Exploration of the bacterial transcriptome**

The advances in transcriptomic analysis are revealing new roles of RNA, and increasing its significance in the original vision of the central dogma proposed by Crick. To date, RNA is not only seen as an information carrier, but is also considered an active player in biological processes. With transcriptomic analysis it is possible to investigate the bacterial transcriptome, highlight genomic features and study the involvement of RNA elements in bacterial physiology and regulation.

So, what does the bacterial transcriptome reveal to us? One initial insight is genome expression. RNA sequences are mirrors of DNA sequences. Therefore, the analysis of RNA population in a microbial cell, allows for determination of which genes are expressed under certain circumstances. Transcriptome analyses directly generate a transcriptional profile at a specific time point during bacterial growth, and allow

investigation of differential gene expression to various stimuli. Changes in gene expression are evidence of metabolic variations and help in understanding the adaptation of the bacterial cell to a new environment [113]. A transcriptome analysis of a bacterial pathogen in the infected tissue can reveal the mechanisms of pathogen-host interaction and clarify the pathogenic process [129].

A primary transcriptome analysis is beneficial for mapping TSSs, improving genome annotation and operon prediction. Identification of TSS consequently allows a deep investigation of 5' UTRs. The 5' UTR of mRNA embraces important functions in gene expression. It includes ribosome binding site (RBS), and non-coding regulatory RNA elements [122], [130]. Moreover, mapping the 5' ends of mRNA forwards the analysis of 5' UTR lengths, the inspection of leader and leaderless genes, and the investigation of promoter regions [122].

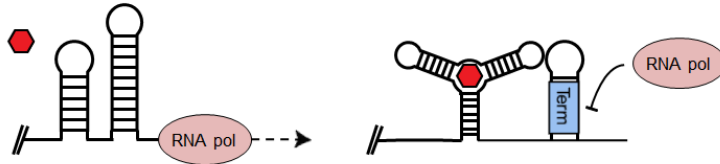
Bacterial whole transcriptome analysis is helpful in discovering non-coding RNA regulators, by highlighting non-coding regions with significant transcriptional activity. Abundant non-coding transcripts have been reported in many bacteria [108], [121], [131]–[134], and are emerging as key players in modulating gene expression by acting through several mechanisms. These RNA molecules are expressed or activated under specific circumstances, and by modulating gene expression, they are responsible for bacterial adaptation to environmental changes and stress conditions (see below) [135].

Finally, transcriptome analysis enables the identification of novel putative ORFs. Indeed, the mapping of transcripts can validate the expression of genes not previously annotated in the genome. This possibility, together with an accurate TSS mapping and operon structure prediction, helps in defining gene structure and genome annotations [96], [97].

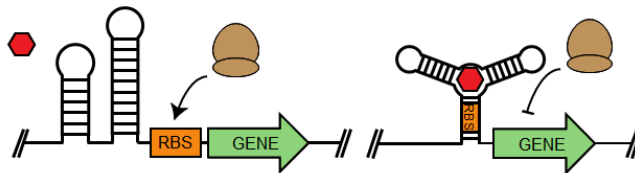
### **2.3.1 Non-coding RNA regulators control gene expression**

The two major classes of RNA regulators are riboswitches and sRNAs [135]. Riboswitches are metabolite-sensing RNA region localized in the 5' UTR of mRNAs. These RNA elements consist of an aptamer and an expression domain. The aptamer recognizes a ligand with high affinity and specificity. The binding of the ligand induces a conformational switch in the expression domain, and the alternative structure alters gene expression. This structural change forms or disrupts a transcriptional terminator or anti-terminator, or blocks or exposes the RBS. Usually, riboswitches repress transcription by forming a terminator. Alternatively, they can act as translation inhibitors by masking the RBS (Figure 10) [136]–[138]. To date, over twenty families of riboswitches with a known ligand have been identified [139], [140]. In addition to these, a number of orphan riboswitches with unknown cognate ligands are also known [138].

#### Transcriptional terminator



#### Translational repressor

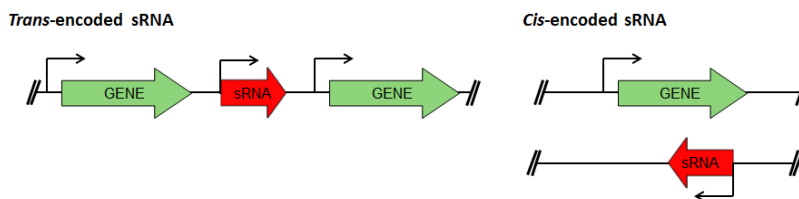


**Figure 10. Riboswitch regulation.** After binding the ligand, riboswitches usually repress gene expression through the formation of transcriptional terminators (top) or by hiding the RBS in the secondary structure (bottom).

A different group of regulatory mechanisms is the one involving sRNAs. Bacterial sRNAs are small (50-500 nucleotide) RNA molecules, which can be encoded between two annotated genes in intergenic regions (*trans*-encoded), or on the opposite strand of an annotated gene (*cis*-encoded) (Figure 11). *Trans*-encoded sRNAs usually regulate multiple mRNA targets distant from their genomic location, by sharing only limited sequence complementarity. In many cases, the help of the RNA chaperone Hfq protein is required to facilitate RNA-RNA interaction. *Cis*-encoded sRNAs are codified on the opposite strand of their target and therefore share a complete sequence complementary with their mRNA. In this case, the RNA-RNA interaction does not require the Hfq



protein [135], [141]. Through base-pairing, sRNAs regulate their mRNA target in many ways: (i) by blocking translation - the pairing on the RBS blocks the translation, and then the sRNA-mRNA duplex is degraded by RNase E [142]–[144], (ii) by activating translation - the pairing overcomes the formation of an inhibitory secondary structure of the target mRNA and activates translation [145], [146], (iii) by increasing mRNA stability - the pairing blocks RNase E recognition or induces cleavage and generate a more stable structure [147], [148]. Besides regulation through RNA base-pairing, a small group of sRNA can also bind proteins [149]–[151].



**Figure 11. *Trans*-encoded and *cis*-encoded sRNAs.** *Trans*-encoded sRNAs are encoded in intergenic region (left), while the *cis*-encoded sRNAs are encoded on the opposite strand of the annotated gene (right).

In summary, transcriptomic approaches have revealed interesting features of bacterial genomes, and display a complex microbial transcriptome, with RNA elements involved in the physiology of prokaryotes. The emerging evidences suggest that there may be more regulatory RNAs with diversified structures and functions that coordinate the bacterial lifestyle.

## 2.4 Transcriptomic analysis of *P. putida* KT2440

With the continuous advances in sequencing techniques, such as RNA-seq and microarrays, several aspects of the *P. putida* KT2440 transcriptome have been investigated. Transcriptome fingerprints of *P. putida* KT2440 growing on different carbon sources, or stress conditions have been generated, and have increased understanding of the metabolic versatility of the bacterium and its adaptation to different environments, as well as revealing non-coding RNA regulatory elements. Since 2006, when the first transcriptomic study (microarray) of *P. putida* KT2440 during the shift from exponential to stationary phase was performed [152], many other transcriptomic studies have followed. A few representative examples are briefly reported here.

In 2007, *del Castillo et al.* clarified the initial step of glucose metabolism in *P. putida* KT2440 by microarray analysis [78]. In 2011, *Frank et al.* refined the annotation of the *P. putida* KT2440 genome by RNA-seq and microarray analysis, and for the first time identified RNA regulatory elements, sRNAs and riboswitches, in KT2440 [153]. *Kim et al.* (2013) and *Nikel et al.* (2014) analyzed the transcriptomic profiles of *P. putida* KT2440 growing on different substrates (glucose, fructose, succinate and glycerol) by RNA-seq [154],[155], while *Sudarsan et al.* (2014) monitored the transcriptomic profiles and the metabolic fluxes of *P. putida* KT2440 shifting from alternative carbon sources (glucose, fructose, and benzoate) [86]. Finally, two transcriptomic studies, Research Articles 2 and 3 of this Ph.D. thesis, have been performed on *P. putida* KT2440 growing under three types of stress conditions and four sole carbon sources, respectively. Research Article 2 has defined the bacterial response to osmotic (NaCl), oxidative (H<sub>2</sub>O<sub>2</sub>), and membrane (Imipenem) stress. Moreover, putative new sRNA candidates have been

identified. Research Article 3 has pointed out the metabolic pathways used by *P. putida* KT2440 growing on glucose, citrate, ferulic acid, and serine, and has identified uptake systems involved in the transport of the carbon sources.

### 3. Concluding remarks and perspectives

The necessity to increase the level knowledge of *P. putida* is essential to develop the bacterium as cell factory, and the information obtained at the transcriptional level by RNA sequencing approaches can reveal interesting bacterial features. Therefore, this Ph.D. work has been built on the overall goal to investigate the transcriptome of *P. putida*, in order to reveal features helpful for the design of the production bacteria. With this in mind, this Ph.D. study has contributed to expand the understanding of *P. putida* KT2440 by three transcriptomic studies.

**Research Article 1** made use of the dRNA-seq approach to map TSSs in *P. putida* KT2440 genome. The approach has allowed further investigation of 5' UTRs and has revealed an interesting high number of TSSs positioned more than 100 or even 200 nucleotides (nt) upstream the first codon. This suggested searching for possible *cis*-acting riboswitch structures in these 5' UTR. The thiamine pyrophosphate (TPP) riboswitch, upstream the *thiC* gene, has been identified and the regulatory mechanism characterized.

**Research Article 2** describes the transcriptome of *P. putida* KT2440 under three stress conditions (NaCl – osmotic; H<sub>2</sub>O<sub>2</sub> – oxidative; Imipenem - membrane), which are often encountered by bacteria during normal growth in the environment and in industrial production. RNA-seq data has highlighted the bacterial stress-response mechanisms after 7 minutes and 1 hour after addition of these stressors to cells growing in exponential phase. A large part of the work has also been focused on the identification and differential expression analysis of new putative sRNAs, which may have relevant roles in controlling the stress response.

Finally, **Research Article 3** shows the capability of *P. putida* KT2440 to grow in the presence of different carbon sources. Here, RNA-seq was

used to describe the metabolic pathways, which allow the bacterium to grow on glucose, citrate, ferulic acid, and serine as sole carbon sources. A metabolic model has been also constructed and flux analyses performed. Particular focus has been placed on the identification of uptake systems involved in the transport of the substrates, and preliminary functional characterization of the transporters has been pursued using mutant strains.

In conclusion, this Ph.D. work has made a significant contribution towards understanding *P. putida* physiology and adaptation to different environments. The results reported here underscore the complexity and diversity of bacterial transcriptomes and provide information helpful for the development of the bacterium as a cell factory and bioremediation tool. The genomic features, metabolic pathways, and RNA elements identified in this Ph.D. work will serve as a benchmark for future studies on *P. putida* gene expression, metabolic engineering, and structural and functional analysis of RNA regulators.

## 4. Bibliography

- [1] V. B. D. Skerman, V. McGowan, and P. H. A. Sneath, "Approved Lists of Bacterial Names," *Med. J. Aust.*, vol. 2, no. 1, pp. 3–4, 1980.
- [2] X. Wu, S. Monchy, S. Taghavi, W. Zhu, J. Ramos, and D. van der Lelie, "Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*," *FEMS Microbiol. Rev.*, vol. 35, no. 2, pp. 299–323, 2011.
- [3] M. A. Matilla, P. Pizarro-Tobias, A. Roca, M. Fernández, E. Duque, L. Molina, X. Wu, D. Van Der Lelie, M. J. Gómez, A. Segura, and J. L. Ramos, "Complete Genome of the Plant Growth-Promoting Rhizobacterium *Pseudomonas putida* BIRD-1," *J. Bacteriol.*, vol. 193, no. 5, p. 1290, 2011.
- [4] J. E. Loper, K. A. Hassan, D. V. Mavrodi, E. W. Davis, C. K. Lim, B. T. Shaffer, L. D. H. Elbourne, V. O. Stockwell, S. L. Hartney, K. Breakwell, M. D. Henkels, S. G. Tetu, L. I. Rangel, T. A. Kidarsa, N. L. Wilson, J. E. van de Mortel, C. Song, R. Blumhagen, D. Radune, J. B. Hostetler, L. M. Brinkac, A. S. Durkin, D. A. Kluepfel, W. P. Wechter, A. J. Anderson, Y. C. Kim, L. S. Pierson, E. A. Pierson, S. E. Lindow, D. Y. Kobayashi, J. M. Raaijmakers, D. M. Weller, L. S. Thomashow, A. E. Allen, and I. T. Paulsen, "Comparative Genomics of Plant-Associated *Pseudomonas* spp.: Insights into Diversity and Inheritance of Traits Involved in Multitrophic Interactions," *PLoS Genet.*, vol. 8, no. 7, 2012.
- [5] D. Balasubramanian, L. Schneper, H. Kumari, and K. Mathee, "A dynamic and intricate regulatory network determines *Pseudomonas aeruginosa* virulence," *Nucleic Acids Res.*, vol. 41, no. 1, pp. 1–20, 2013.
- [6] A. Rodríguez-Rojas, A. Oliver, and J. Blázquez, "Intrinsic and Environmental Mutagenesis Drive Diversification and Persistence of *Pseudomonas aeruginosa* in Chronic Lung Infections," *J. Infect. Dis.*, vol. 205, no. 1, pp. 121–127, 2011.
- [7] T. B. May and A. M. Chakrabarty, "*Pseudomonas aeruginosa*: genes and enzymes of alginate synthesis," *Trends Microbiol.*, vol. 2, no. 5, pp. 151–157, 1994.
- [8] I. Poblete-Castro, J. Becker, K. Dohnt, V. M. dos Santos, and C. Wittmann, "Industrial biotechnology of *Pseudomonas putida* and related species," *Appl. Microbiol. Biotechnol.*, vol. 93, no. 6, pp. 2279–2290, 2012.
- [9] P. I. Nikel, E. Martínez-García, and V. de Lorenzo, "Biotechnological

- domestication of pseudomonads using synthetic biology,” *Nat. Rev. Microbiol.*, vol. 12, no. 5, pp. 368–379, 2014.
- [10] T. Tiso, N. Wierckx, and L. Blank, "Non-Pathogenic *Pseudomonas* as Platform for Industrial Biocatalysis", vol. 1. 2014.
  - [11] K. E. Nelson, C. Weinel, I. T. Paulsen, R. J. Dodson, H. Hilbert, V. A. P. Martins dos Santos, D. E. Fouts, S. R. Gill, M. Pop, M. Holmes, V. A. P. M. Santos, D. E. Fouts, S. R. Gill, M. Pop, M. Holmes, L. Brinkac, M. Beanan, R. T. Deboy, S. Daugherty, J. Kolonay, R. Madupu, W. Nelson, O. White, J. Peterson, H. Khouri, I. Hance, P. C. Lee, E. Holtzapple, D. Scanlan, K. Tran, A. Moazzez, T. Utterback, M. Rizzo, K. Lee, D. Kosack, D. Moestl, H. Wedler, J. Lauber, D. Stjepandic, J. Hoheisel, M. Straetz, S. Heim, C. Kiewitz, J. Eisen, K. N. Timmis, A. D sterh ft, B. T mmler, and C. M. Fraser, "Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440," *Environ. Microbiol.*, vol. 4, no. 12, pp. 799–808, 2002.
  - [12] K. N. Timmis, "*Pseudomonas putida*: a cosmopolitan opportunist par excellence," *Environ. Microbiol.*, vol. 4, no. 12, pp. 779–781, 2002.
  - [13] V. A. P. Martins dos Santos, S. Heim, E. R. B. Moore, M. Str tz, and K. N. Timmis, "Insights into the genomic basis of niche specificity of *Pseudomonas putida* KT2440," *Environ. Microbiol.*, vol. 6, no. 12, pp. 1264–1286, 2004.
  - [14] B. J. J. Lugtenberg, T. F. C. Chin-a-woeng, and G. V Bloemberg, "Microbe-plant interactions: principles and mechanisms," *Antonie Van Leeuwenhoek*, vol. 81, no. 1–4, pp. 373–383, 2002.
  - [15] F. Persello-Cartieaux, L. Nussaume, and C. Robaglia, "Tales from the underground: molecular plant-rhizobacteria interactions," *Plant, Cell Environ.*, vol. 26, no. 2, pp. 189–199, 2003.
  - [16] R. La Rosa, J. Nogales, and F. Rojo, "The Crc/CrcZ-CrcY global regulatory system helps the integration of gluconeogenic and glycolytic metabolism in *Pseudomonas putida*," *Environ. Microbiol.*, vol. 17, no. 9, pp. 3362–3378, 2015.
  - [17] F. Rojo, "Carbon catabolite repression in *Pseudomonas*: optimizing metabolic versatility and interactions with the environment," *FEMS Microbiol. Rev.*, vol. 34, no. 5, pp. 658–684, 2010.
  - [18] J. A. Wolff, C. H. MacGregor, R. C. Eisenberg, and P. V. Phibbs, "Isolation and Characterization of Catabolite Repression Control Mutants of *Pseudomonas aeruginosa* PAO," *J. Bacteriol.*, vol. 173, no. 15, pp. 4700–4706, 1991.
  - [19] R. La Rosa, V. Behrends, H. D. Williams, J. G. Bundy, and F. Rojo, "Influence of the Crc regulator on the hierarchical use of carbon sources

- from a complete medium in *Pseudomonas*,” *Environ. Microbiol.*, vol. 18, no. 3, pp. 807–18, 2015.
- [20] R. Moreno, M. Martínez-Gomariz, L. Yuste, C. Gil, and F. Rojo, “The *Pseudomonas putida* Crc global regulator controls the hierarchical assimilation of amino acids in a complete medium: Evidence from proteomic and genomic analyses,” *Proteomics*, vol. 9, no. 11, pp. 2910–2928, 2009.
  - [21] R. Y. Stanier, N. J. Palleroni, and M. Doudoroff, “The Aerobic *Pseudomonads*: A Taxonomic Study,” *J. Gen. Microbiol.*, vol. 43, no. 2, pp. 159–271, 1966.
  - [22] A. Schmid, J. S. Dordick, B. Hauer, A. Kiener, M. Wubbolts, and B. Witholt, “Industrial biocatalysis today and tomorrow,” *Nature*, vol. 409, no. 6817, pp. 258–268, 2001.
  - [23] J. I. Jiménez, B. Miñambres, J. L. García, and E. Díaz, “Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440,” *Environ. Microbiol.*, vol. 4, no. 12, pp. 824–841, 2002.
  - [24] G. Morales, J. F. Linares, A. Beloso, J. P. Albar, and F. Rojo, “The *Pseudomonas putida* Crc Global Regulator Controls the Expression of Genes from Several Chromosomal Catabolic Pathways for Aromatic Compounds,” *J. Bacteriol.*, vol. 186, no. 5, pp. 1337–1344, 2004.
  - [25] R. Moreno and F. Rojo, “The Target for the *Pseudomonas putida* Crc Global Regulator in the Benzoate Degradation Pathway Is the BenR Transcriptional Regulator,” *J. Bacteriol.*, vol. 190, no. 5, pp. 1539–1545, 2008.
  - [26] O. N. Reva, C. Weinel, M. Weinel, K. Böhm, D. Stjepandic, J. D. Hoheisel, and B. Tümmeler, “Functional genomics of stress response in *Pseudomonas putida* KT2440,” *J. Bacteriol.*, vol. 188, no. 11, pp. 4079–92, 2006.
  - [27] P. Fonseca, R. Moreno, and F. Rojo, “Growth of *Pseudomonas putida* at low temperature: global transcriptomic and proteomic analyses,” *Environ. Microbiol. Rep.*, vol. 3, no. 3, pp. 329–339, 2011.
  - [28] M. Chavarría, P. I. Nikel, D. Pérez-Pantoja, and V. De Lorenzo, “The Entner-Doudoroff pathway empowers *Pseudomonas putida* KT2440 with a high tolerance to oxidative stress,” *Environ. Microbiol.*, vol. 15, no. 6, pp. 1772–1785, 2013.
  - [29] S. Hishinuma, M. Yuki, M. Fujimura, and F. Fukumori, “OxyR regulated the expression of two major catalases, KatA and KatB, along with peroxiredoxin, AhpC in *Pseudomonas putida*,” *Environ. Microbiol.*, vol. 8, no. 12, pp. 2115–2124, 2006.
  - [30] J. Kim, C. O. Jeon, and W. Park, “Dual regulation of zwf-1 by both 2-keto-3-deoxy-6-phosphogluconate and oxidative stress in *Pseudomonas*



- putida*,” *Microbiology*, vol. 154, no. 12, pp. 3905–3916, 2008.
- [31] W. Park, S. Peña-Llopis, Y. Lee, and B. Dimple, “Regulation of superoxide stress in *Pseudomonas putida* KT2440 is different from the SoxR paradigm in *Escherichia coli*,” *Biochem. Biophys. Res. Commun.*, vol. 341, no. 1, pp. 51–56, 2006.
  - [32] J. Yeom, C. O. Jeon, E. L. Madsen, and W. Park, “Ferredoxin-NADP+ Reductase from *Pseudomonas putida* Functions as a Ferric Reductase,” *J. Bacteriol.*, vol. 191, no. 5, pp. 1472–1479, 2009.
  - [33] J. Yeom, C. O. Jeon, E. L. Madsen, and W. Park, “In vitro and In vivo Interactions of Ferredoxin-NADP+ Reductases in *Pseudomonas putida*,” *J. Biochem.*, vol. 145, no. 4, pp. 481–491, 2009.
  - [34] S. Yeom, J. Yeom, and W. Park, “NtrC-Sensed Nitrogen Availability Is Important for Oxidative Stress Defense in *Pseudomonas putida* KT2440,” *J. Microbiol.*, vol. 48, no. 2, pp. 153–159, 2010.
  - [35] S. Yeom, J. Yeom, and W. Park, “Molecular characterization of FinR, a novel redox-sensing transcriptional regulator in *Pseudomonas putida* KT2440,” *Microbiology*, vol. 156, no. 5, pp. 1487–1496, 2010.
  - [36] J. Yeom, Y. Lee, and W. Park, “ATP-dependent RecG Helicase Is Required for the Transcriptional Regulator OxyR Function in *Pseudomonas* species,” *J. Biol. Chem.*, vol. 287, no. 29, pp. 24492–24504, 2012.
  - [37] V. Kumar Deshwal and P. Kumar, “Effect of Salinity on Growth and PGPR Activity of Pseudomonads,” *J. Acad. Ind. Res.*, vol. 2, no. 6, pp. 353–356, 2013.
  - [38] A. G. De Castro, H. Bredholt, A. R. Strøm, and A. Tunnacliffe, “Anhydrobiotic Engineering of Gram-Negative Bacteria,” *Appl. Environ. Microbiol.*, vol. 66, no. 9, pp. 4142–4144, 2000.
  - [39] M. Manzanera, A. García de Castro, A. Tøndervik, M. Rayner-Brandes, A. R. Strøm, and A. Tunnacliffe, “Hydroxyectoine Is Superior to Trehalose for Anhydrobiotic Engineering of *Pseudomonas putida* KT2440,” *Appl. Environ. Microbiol.*, vol. 68, no. 9, pp. 4328–4333, 2002.
  - [40] J. E. Hallsworth, S. Heim, and K. N. Timmis, “Chaotropic solutes cause water stress in *Pseudomonas putida*,” *Environ. Microbiol.*, vol. 5, no. 12, pp. 1270–1280, 2003.
  - [41] T. Nakazawa and T. Yokota, “Benzoate Metabolism in *Pseudomonas putida*(arvilla) mt-2: Demonstration of Two Benzoate Pathways,” *J. Bacteriol.*, vol. 115, no. 1, pp. 262–267, 1973.
  - [42] N. W. Dunn and I. C. Gunsalus, “Transmissible Plasmid Coding Early Enzymes of Naphthalene Oxidation in *Pseudomonas putida*,” *J. Bacteriol.*, vol. 114, no. 3, pp. 974–979, 1973.

- [43] D. Zhen, H. Liu, S. J. Wang, J. J. Zhang, F. Zhao, and N. Y. Zhou, "Plasmid-mediated degradation of 4-chloronitrobenzene by newly isolated *Pseudomonas putida* strain ZWL73," *Appl. Microbiol. Biotechnol.*, vol. 72, no. 4, pp. 797–803, 2006.
- [44] H. F. Dean, S. Cheevadhanarak, R. A. Skurray, and R. C. Bayly, "Characterisation of a degradative plasmid in *Pseudomonas putida* that controls the expression of 2, 4-xyleneol degradative genes," *FEMS Microbiol. Lett.*, vol. 61, pp. 153–157, 1989.
- [45] H. Herrmann, D. Janke, S. Krejsa, and I. Kunze, "Involvement of the plasmid pPGH1 in the phenol degradation of *Pseudomonas putida* strain H," *FEMS Microbiol. Lett.*, vol. 43, pp. 133–137, 1987.
- [46] J. L. Ramos, M. S. Cuenca, C. Molina-Santiago, A. Segura, E. Duque, M. R. Gomez-García, Z. Udaondo, and A. Roca, "Mechanisms of solvent resistance mediated by interplay of cellular factors in *Pseudomonas putida*," *FEMS Microbiol. Rev.*, vol. 39, no. 4, pp. 555–566, 2015.
- [47] A. Loeschcke and S. Thies, "*Pseudomonas putida*—a versatile host for the production of natural products," *Appl. Microbiol. Biotechnol.*, vol. 99, no. 15, pp. 6197–6214, 2015.
- [48] N. Hoffmann and B. H. A. Rehm, "Regulation of polyhydroxyalkanoate biosynthesis in *Pseudomonas putida* and *Pseudomonas aeruginosa*," *FEMS Microbiol. Lett.*, vol. 237, no. 1, pp. 1–7, 2004.
- [49] S. Khanna and A. K. Srivastava, "Recent advances in microbial polyhydroxyalkanoates," *Process Biochem.*, vol. 40, no. 2, pp. 607–619, 2005.
- [50] A. Steinbüchel and H. E. Valentin, "Diversity of bacterial polyhydroxyalkanoic acids," *FEMS Microbiol. Lett.*, vol. 128, pp. 219–228, 1995.
- [51] S. J. Park, T. W. Kim, M. K. Kim, S. Y. Lee, and S. C. Lim, "Advanced bacterial polyhydroxyalkanoates: Towards a versatile and sustainable platform for unnatural tailor-made polyesters," *Biotechnol. Adv.*, vol. 30, no. 6, pp. 1196–1206, 2012.
- [52] P. Fonseca, F. de la Peña, and M. A. Prieto, "A role for the regulator PsrA in the polyhydroxyalkanoate metabolism of *Pseudomonas putida* KT2440," *Int. J. Biol. Macromol.*, vol. 71, pp. 14–20, 2014.
- [53] P. I. Nikel, A. De Almeida, E. C. Melillo, M. A. Galvagno, and M. J. Pettinari, "New Recombinant *Escherichia coli* Strain Tailored for the Production of Poly(3-Hydroxybutyrate) from Agroindustrial By-Products," *Appl. Environ. Microbiol.*, vol. 72, no. 6, pp. 3949–3954, 2006.
- [54] Z. Sun, J. A. Ramsay, M. Guay, and B. A. Ramsay, "Carbon-limited

- fed-batch production of medium-chain-length polyhydroxyalkanoates from nonanoic acid by *Pseudomonas putida* KT2440,” *Appl. Microbiol. Biotechnol.*, vol. 82, no. 4, pp. 657–662, 2009.
- [55] W. Liu and G. Q. Chen, “Production and characterization of medium-chain-length polyhydroxyalkanoate with high 3-hydroxytetradecanoate monomer content by *fadB* and *fadA* knockout mutant of *Pseudomonas putida* KT2442,” *Appl. Microbiol. Biotechnol.*, vol. 76, no. 5, pp. 1153–1159, 2007.
  - [56] X. Gao, J. C. Chen, Q. Wu, and G. Q. Chen, “Polyhydroxyalkanoates as a source of chemicals, polymers, and biofuels,” *Curr. Opin. Biotechnol.*, vol. 22, no. 6, pp. 768–774, 2011.
  - [57] J. B. J. H. Van Duuren, D. Wijte, B. Karge, V. A. P. Martins dos Santos, Y. Yang, A. E. Mars, and G. Eggink, “PH-Stat Fed-Batch Process to Enhance the Production of *cis*, *cis*-Muconate from Benzoate by *Pseudomonas putida* KT2440-JD1,” *Biotechnol. Prog.*, vol. 28, no. 1, pp. 85–92, 2012.
  - [58] R. Plaggenborg, J. Overhage, a Steinbüchel, and H. Priefert, “Functional analyses of genes involved in the metabolism of ferulic acid in *Pseudomonas putida* KT2440,” *Appl. Microbiol. Biotechnol.*, vol. 61, no. 5–6, pp. 528–535, 2003.
  - [59] N. Graf and J. Altenbuchner, “Genetic engineering of *Pseudomonas putida* KT2440 for rapid and high-yield production of vanillin from ferulic acid,” *Appl. Microbiol. Biotechnol.*, vol. 98, no. 1, pp. 137–149, 2014.
  - [60] P. G. Ward, M. Goff, M. Donner, W. Kaminsky, and K. E. O’Connor, “A Two Step Chemo-biotechnological Conversion of Polystyrene to a Biodegradable Thermoplastic,” *Environ. Sci. Technol.*, vol. 40, no. 7, pp. 2433–2437, 2006.
  - [61] K. Hahlbrock and D. Scheel, “Physiology and molecular biology of phenylpropanoid metabolism,” *Plant Mol. Biol.*, vol. 40, pp. 347–369, 1989.
  - [62] M. Matsusaki, H. T. Tran, T. Kaneko, and M. Akashi, “Enhanced effects of lithocholic acid incorporation into liquid-crystalline biopolymer poly(coumaric acid) on structural ordering and cell adhesion,” *Biomaterials*, vol. 26, no. 32, pp. 6263–6270, 2005.
  - [63] R. Benkrief, Y. Ranarivelo, A. L. Skaltsounis, F. Tillequin, M. Koch, J. Pusset, and T. Sévenet, “Monoterpene alkaloids, iridoids and phenylpropanoid glycosides from *Osmanthus austrocaledonica*,” *Phytochemistry*, vol. 47, no. 5, pp. 825–832, 1998.
  - [64] K. Nijkamp, R. G. M. Westerhof, H. Ballerstedt, J. A. M. De Bont, and J. Wery, “Optimization of the solvent-tolerant *Pseudomonas putida* S12

- as host for the production of p-coumarate from glucose,” *Appl. Microbiol. Biotechnol.*, vol. 74, no. 3, pp. 617–624, 2007.
- [65] K. Gerth, S. Pradella, O. Perlova, S. Beyer, and R. Müller, “Myxobacteria: proficient producers of novel natural products with various biological activities-past and future biotechnological aspects with the focus on the genus *Sorangium*,” *J. Biotechnol.*, vol. 106, no. 2–3, pp. 233–253, 2003.
  - [66] S. Stephan, E. Heinzle, S. C. Wenzel, D. Krug, R. Müller, and C. Wittmann, “Metabolic physiology of *Pseudomonas putida* for heterologous production of myxochromide,” *Process Biochem.*, vol. 41, no. 10, pp. 2146–2152, 2006.
  - [67] C. Choi, E. C. Park, S. H. Yun, S. Lee, Y. G. Lee, Y. Hong, K. R. Park, S. Kim, G. Kim, and S. Il Kim, “Proteomic Characterization of the Outer Membrane Vesicle of *Pseudomonas putida* KT2440,” *J. Proteome Res.*, vol. 13, no. 10, pp. 4298–4309, 2014.
  - [68] B. R. Glick, “Plant Growth-Promoting Bacteria: Mechanisms and Applications,” *Scientifica (Cairo)*, vol. 2012, p. 963401, 2012.
  - [69] D. Regenhardt, H. Heuer, S. Heim, D. U. Fernandez, C. Strömpl, E. R. B. Moore, and K. N. Timmis, “Pedigree and taxonomic credentials of *Pseudomonas putida* strain KT2440,” *Environ. Microbiol.*, vol. 4, no. 12, pp. 912–915, 2002.
  - [70] E. Belda, R. G. A. Van Heck, M. J. Lopez-sanchez, and S. Cruveiller, “The revisited genome of *Pseudomonas putida* KT2440 enlightens its value as a robust metabolic chassis,” *Environ. Microbiol.*, 2016.
  - [71] T. Nakazawa, “Travels of a *Pseudomonas*, from Japan around the world,” *Environ. Microbiol.*, vol. 4, no. 12, pp. 782–786, 2002.
  - [72] P. Godoy, A. J. Molina-Henares, J. De La Torre, E. Duque, and J. L. Ramos, “Characterization of the RND family of multidrug efflux pumps: in silico to in vivo confirmation of four functionally distinct subgroups,” *Microb. Biotechnol.*, vol. 3, no. 6, pp. 691–700, 2010.
  - [73] E. Martínez-García and V. de Lorenzo, “Engineering multiple genomic deletions in Gram-negative bacteria: analysis of the multi-resistant antibiotic profile of *Pseudomonas putida* KT2440,” *Environ. Microbiol.*, vol. 13, no. 10, pp. 2702–2716, 2011.
  - [74] M. Fernández, S. Conde, J. de la Torre, C. Molina-Santiago, J.-L. Ramos, and E. Duque, “Mechanisms of Resistance to Chloramphenicol in *Pseudomonas putida* KT2440,” *Antimicrob. Agents Chemother.*, vol. 56, no. 2, pp. 1001–1009, 2012.
  - [75] C. Roma-Rodrigues, P. M. Santos, D. Benndorf, E. Rapp, and I. Sá-Correia, “Response of *Pseudomonas putida* KT2440 to phenol at the level of membrane proteome,” *J. Proteomics*, vol. 73, no. 8, pp. 1461–

- 1478, 2010.
- [76] A. Roca, J. J. Rodríguez-Herva, E. Duque, and J. L. Ramos, "Physiological responses of *Pseudomonas putida* to formaldehyde during detoxification," *Microb. Biotechnol.*, vol. 1, no. 2, pp. 158–169, 2008.
  - [77] P. Domínguez-Cuevas, J. E. González-Pastor, S. Marqués, J. L. Ramos, and V. De Lorenzo, "Transcriptional tradeoff between metabolic and stress-response programs in *Pseudomonas putida* KT2440 cells exposed to toluene," *J. Biol. Chem.*, vol. 281, no. 17, pp. 11981–11991, 2006.
  - [78] T. Del Castillo, J. L. Ramos, J. J. Rodríguez-Herva, T. Fuhrer, U. Sauer, and E. Duque, "Convergent Peripheral Pathways Catalyze Initial Glucose Catabolism in *Pseudomonas putida*: Genomic and Flux Analysis," *J. Bacteriol.*, vol. 189, no. 14, pp. 5142–5152, 2007.
  - [79] J. Nogales, B. Ø. Palsson, and I. Thiele, "A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iJN746 as a cell factory," *BMC Syst. Biol.*, vol. 2, no. 1, p. 79, 2008.
  - [80] J. Puchalka, M. A. Oberhardt, M. Godinho, A. Bielecka, D. Regenhart, K. N. Timmis, J. A. Papin, and V. A. P. Martins Dos Santos, "Genome-Scale Reconstruction and Analysis of the *Pseudomonas putida* KT2440 Metabolic Network Facilitates Applications in Biotechnology," *PLoS Comput. Biol.*, vol. 4, no. 10, 2008.
  - [81] F. Velázquez, K. Pflüger, I. Cases, L. I. De Eugenio, and V. De Lorenzo, "The Phosphotransferase System Formed by PtsP, PtsO, and PtsN Proteins Controls Production of Polyhydroxyalkanoates in *Pseudomonas putida*," *J. Bacteriol.*, vol. 189, no. 12, pp. 4529–4533, 2007.
  - [82] N. Entner and M. Doudoroff, "Glucose and gluconic acid oxidation of *Pseudomonas saccharophila*," *J Biol Chem*, vol. 196, no. 2, pp. 853–862, 1952.
  - [83] T. Fuhrer, E. Fischer, and U. Sauer, "Experimental Identification and Quantification of Glucose Metabolism in Seven Bacterial Species," *J. Bacteriol.*, vol. 187, no. 5, pp. 1581–1590, 2005.
  - [84] E. G. Saravolac, N. F. Taylor, R. Benz, and R. E. W. Hancock, "Purification of Glucose-Inducible Outer Membrane Protein OprB of *Pseudomonas putida* and Reconstitution of Glucose-Specific Pores," *J. Bacteriol.*, vol. 173, no. 16, pp. 4970–4976, 1991.
  - [85] J. L. Wylie and E. A. Worobec, "The OprB Porin Plays a Central Role in Carbohydrate Uptake in *Pseudomonas aeruginosa*," *J. Bacteriol.*, vol. 177, no. 11, pp. 3021–3026, 1995.
  - [86] S. Sudarsan, S. Dethlefsen, L. M. Blank, M. Siemann-Herzberg, and A. Schmid, "The Functional Structure of Central Carbon Metabolism in

- Pseudomonas putida* KT2440,” *Appl. Environ. Microbiol.*, vol. 80, no. 17, pp. 5292–5303, 2014.
- [87] Y. H. Kim, K. Cho, S. H. Yun, J. Y. Kim, K. H. Kwon, J. S. Yoo, and S. Il Kim, “Analysis of aromatic catabolic pathways in *Pseudomonas putida* KT2440 using a combined proteomic approach: 2-DE/MS and cleavable isotope-coded affinity tag analysis,” *Proteomics*, vol. 6, no. 4, pp. 1301–1318, 2006.
  - [88] S. B. Sohn, T. Y. Kim, J. M. Park, and S. Y. Lee, “In silico genome-scale metabolic analysis of *Pseudomonas putida* KT2440 for polyhydroxyalkanoate synthesis, degradation of aromatics and anaerobic survival,” *Biotechnol. J.*, vol. 5, no. 7, pp. 739–750, 2010.
  - [89] M. A. Oberhardt, J. Puchałka, V. A. P. M. dos Santos, and J. A. Papin, “Reconciliation of Genome-Scale Metabolic Reconstructions for Comparative Systems Analysis,” *PLoS Comput. Biol.*, vol. 7, no. 3, 2011.
  - [90] F. Crick, “Central Dogma of Molecular Biology,” *Nature*, vol. 227, no. 5258, pp. 561–563, 1970.
  - [91] S. Franklin and T. M. Vondriska, “Genomes, Proteomes, and the Central Dogma,” *Circ. Cardiovasc. Genet.*, vol. 4, no. 5, p. 576, 2011.
  - [92] V. De Lorenzo, “From the selfish gene to selfish metabolism: Revisiting the central dogma,” *BioEssays*, vol. 36, no. 3, pp. 226–235, 2014.
  - [93] A. D. Ellington, “RNA as a conception,” *Rna*, vol. 21, no. 4, pp. 608–608, 2015.
  - [94] K. Bandyra and B. Luisi, “Central dogma alchemy,” *RNA*, vol. 21, no. 4, pp. 558–559, 2015.
  - [95] K. V. Morris and M. J.S., “The rise of regulatory RNA,” *Nat. Rev. Genet.*, vol. 15, no. 6, pp. 423–437, 2014.
  - [96] R. Sorek and P. Cossart, “Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity,” *Nat Rev Genet*, vol. 11, no. 1, pp. 9–16, 2010.
  - [97] M. Güell, E. Yus, M. Lluch-Senar, and L. Serrano, “Bacterial transcriptomics: what is beyond the RNA horiz-ome?,” *Nat. Rev. Microbiol.*, vol. 9, no. 9, pp. 658–669, 2011.
  - [98] F. Sanger, S. Nicklen, and A. R. Coulson, “DNA sequencing with chain-terminating inhibitors,” *Proc Natl Acad Sci USA*, vol. 74, no. 12, pp. 5463–5467, 1977.
  - [99] E. R. Mardis, “Next-Generation Sequencing Platforms,” *Annu. Rev. Anal. Chem.*, vol. 6, no. 1, pp. 287–303, 2013.
  - [100] G. J. Bauer, “RNA sequencing using fluorescent-labeled dideoxynucleotides and automated fluorescence detection,” *Nucleic Acids Res.*, vol. 18, no. 4, pp. 879–884, 1990.

- [101] C. A. Hutchison, "DNA sequencing: bench to bedside and beyond," *Nucleic Acids Res.*, vol. 35, no. 18, pp. 6227–6237, 2007.
- [102] C. J. Davidson, E. Zeringer, K. J. Champion, M. P. Gauthier, F. Wang, J. Boonyaratanakornkit, J. R. Jones, and E. Schreiber, "Improving the limit of detection for Sanger sequencing: A comparison of methodologies for KRAS variant detection," *Biotechniques*, vol. 53, no. 3, pp. 182–188, 2012.
- [103] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown, "Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray," *Science (New York, N.Y.)*, vol. 270, no. 5235, pp. 467–470, 1995.
- [104] X. S. Liu, "Getting Started in Tiling Microarray Analysis," *PLoS Comput. Biol.*, vol. 3, no. 10, pp. 1842–1844, 2007.
- [105] T. C. Mockler and J. R. Ecker, "Applications of DNA tiling arrays for whole-genome analysis," *Genomics*, vol. 85, no. 1, pp. 1–15, 2005.
- [106] D. W. Selinger, K. J. Cheung, R. Mei, E. M. Johansson, C. S. Richmond, F. R. Blattner, D. J. Lockhart, and G. M. Church, "RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array," *Nat. Biotechnol.*, vol. 18, no. 12, pp. 1262–1268, 2000.
- [107] P. T. McGrath, H. Lee, L. Zhang, A. A. Iniesta, A. K. Hottes, M. H. Tan, N. J. Hillson, P. Hu, L. Shapiro, and H. H. McAdams, "High-throughput identification of transcription start sites, conserved promoter motifs and predicted regulons," *Nat. Biotechnol.*, vol. 25, no. 5, pp. 584–592, 2007.
- [108] A. Toledo-Arana, O. Dussurget, G. Nikitas, N. Sesto, H. Guet-Revillet, D. Balestrino, E. Loh, J. Gripenland, T. Tiensuu, K. Vaitkevicius, M. Barthelemy, M. Vergassola, M.-A. Nahori, G. Soubigou, B. Régnault, J.-Y. Coppée, M. Lecuit, J. Johansson, and P. Cossart, "The *Listeria* transcriptional landscape from saprophytism to virulence," *Nature*, vol. 459, no. 7249, pp. 950–956, 2009.
- [109] S. Rasmussen, H. B. Nielsen, and H. Jarmer, "The transcriptionally active regions in the genome of *Bacillus subtilis*," *Mol. Microbiol.*, vol. 73, no. 6, pp. 1043–1057, 2009.
- [110] J. C. D. Hinton, I. Hautefort, S. Eriksson, A. Thompson, and M. Rhen, "Benefits and pitfalls of using microarrays to monitor bacterial gene expression during infection," *Curr. Opin. Microbiol.*, vol. 7, no. 3, pp. 277–282, 2004.
- [111] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics," *Nat. Rev. Genet.*, vol. 10, no. 1, pp. 57–63, 2009.
- [112] A. Roberts, C. Trapnell, J. Donaghey, J. L. Rinn, and L. Pachter, "Improving RNA-Seq expression estimates by correcting for fragment

- bias,” *Genome Biol.*, vol. 12, no. 3, p. R22, 2011.
- [113] M. J. Filiatrault, “Progress in prokaryotic transcriptomics,” *Curr. Opin. Microbiol.*, vol. 14, no. 5, pp. 579–586, Oct. 2011.
  - [114] J. Z. Levin, M. Yassour, X. Adiconis, C. Nusbaun, D. A. Thompson, N. Friedman, A. Gnirke, and A. Regev, “Comprehensive comparative analysis of strand specific RNA sequencing methods,” *Nat. Methods*, vol. 7, no. 9, pp. 709–715, 2010.
  - [115] J. D. Mills, Y. Kawahara, and M. Janitz, “Strand-Specific RNA-Seq Provides Greater Resolution of Transcriptome Profiling,” *Curr. Genomics*, vol. 14, no. 3, pp. 173–81, 2013.
  - [116] D. Parkhomchuk, T. Borodina, V. Amstislavskiy, M. Banaru, L. Hallen, S. Krobitsch, H. Lehrach, and A. Soldatov, “Transcriptome analysis by strand-specific sequencing of complementary DNA,” *Nucleic Acids Res.*, vol. 37, no. 18, 2009.
  - [117] L. Wang, Y. Si, L. K. Dedow, Y. Shao, P. Liu, and T. P. Brutnell, “A Low-Cost Library Construction Protocol and Data Analysis Pipeline for Illumina-Based Strand-Specific Multiplex RNA-seq,” *PLoS One*, vol. 6, no. 10, 2011.
  - [118] T. Borodina, J. Adjaye, and M. Sultan, “A Strand-Specific Library Preparation Protocol for RNA Sequencing,” 1st ed., vol. 500. *Elsevier Inc.*, 2011.
  - [119] M. Sultan, S. Dökel, V. Amstislavskiy, D. Wuttig, H. Sülthmann, H. Lehrach, and M. L. Yaspo, “A simple strand-specific RNA-Seq library preparation protocol combining the Illumina TruSeq RNA and the dUTP methods,” *Biochem. Biophys. Res. Commun.*, vol. 422, no. 4, pp. 643–646, 2012.
  - [120] Z. Zhang, W. E. Theurkauf, Z. Weng, and P. D. Zamore, “Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection,” *Silence*, vol. 3, no. 1, p. 9, 2012.
  - [121] C. M. Sharma, S. Hoffmann, F. Darfeuille, J. Reignier, S. Findeiß, A. Sittka, S. Chabas, K. Reiche, J. Hackermüller, R. Reinhardt, P. F. Stadler, and J. Vogel, “The primary transcriptome of the major human pathogen *Helicobacter pylori*,” *Nature*, vol. 464, no. 7286, pp. 250–255, 2010.
  - [122] C. M. Sharma and J. Vogel, “Differential RNA-seq: the approach behind and the biological insight gained,” *Curr. Opin. Microbiol.*, vol. 19C, pp. 97–105, 2014.
  - [123] G. Dugar, A. Herbig, K. U. Förstner, N. Heidrich, R. Reinhardt, K. Nieselt, and C. M. Sharma, “High-Resolution Transcriptome Maps Reveal Strain-Specific Regulatory Features of Multiple *Campylobacter jejuni* Isolates,” *PLoS Genet.*, vol. 9, no. 5, p. e1003495, 2013.



- [124] I. Irtsov, C. M. Sharma, J. Vogel, and W. C. Winkler, "Identification of regulatory RNAs in *Bacillus subtilis*," *Nucleic Acids Res.*, vol. 38, no. 19, pp. 6637–51, 2010.
- [125] M. K. Thomason, T. Bischler, S. K. Eisenbart, K. U. Förstner, A. Zhang, A. Herbig, K. Nieselt, C. M. Sharma, and G. Storz, "Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*," *J. Bacteriol.*, vol. 197, no. 1, pp. 18–28, 2015.
- [126] T. Cortes, O. T. Schubert, G. Rose, K. B. Arnvig, I. Comas, R. Aebersold, and D. B. Young, "Genome-wide Mapping of Transcriptional Start Sites Defines an Extensive Leaderless Transcriptome in *Mycobacterium tuberculosis*," *Cell Rep.*, vol. 5, no. 4, pp. 1121–1131, 2013.
- [127] M. J. Filiatrault, P. V. Stodghill, C. R. Myers, P. A. Bronstein, B. G. Butcher, H. Lam, G. Grills, P. Schweitzer, W. Wang, D. J. Schneider, and S. W. Cartinhour, "Genome-Wide Identification of Transcriptional Start Sites in the Plant Pathogen *Pseudomonas syringae* pv. *tomato* str. dc3000," *PLoS One*, vol. 6, no. 12, p. e29335, 2011.
- [128] C. Kröger, S. Dillon, A. Cameron, K. Papenfort, S. Sivasankaran, K. Hokamp, Y. Chao, A. Sittka, M. Hébrard, K. Händler, A. Colgan, P. Leekitcharoenphon, G. Langridge, A. Lohan, B. Loftus, S. Lucchini, D. Ussery, C. Dorman, N. Thomson, J. Vogel, and J. Hinton, "The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 20, pp. 1277–1286, 2012.
- [129] T. a. Skvortsov and T. L. Azhikina, "A Review of the Transcriptome Analysis of Bacterial Pathogens in vivo: Problems and Solutions," *Russ. J. Bioorganic Chem.*, vol. 36, no. 5, pp. 550–559, 2010.
- [130] S. Cho, Y. Cho, S. Lee, J. Kim, H. Yum, S. C. Kim, and B.-K. Cho, "Current Challenges in Bacterial Transcriptomics," *Genomics Inform.*, vol. 11, no. 2, pp. 76–82, 2013.
- [131] M. Guell van Noort, V., Yus, E., Chen, W. H., Leigh-Bell, J., Michalodimitrakis, K., Yamada, T., Arumugam, M., Doerks, T., Kuhner, S., Rode, M., Suyama, M., Schmidt, S., Gavin, A. C., Bork, P., Serrano, L., "Transcriptome Complexity in a Genome-Reduced Bacterium," *Science (80- )*, vol. 326, no. 5957, pp. 1268–1271, 2009.
- [132] D. Yoder-Himes, P. Chain, Y. Zhu, E. Rubin, O. Wurtzel, J. Tiedje, and R. Sorek, "Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing," *Proc Natl Acad Sci USA*, vol. 106, no. 10, pp. 3976–3981, 2009.
- [133] T. T. Perkins, R. A. Kingsley, M. C. Fookes, P. P. Gardner, K. D.

- James, L. Yu, S. A. Assefa, M. He, N. J. Croucher, D. J. Pickard, D. J. Maskell, J. Parkhill, J. Choudhary, N. R. Thomson, and G. Dougan, "A Strand-Specific RNA-Seq Analysis of the Transcriptome of the Typhoid *Bacillus Salmonella* Typhi," *PLoS Genet.*, vol. 5, no. 7, 2009.
- [134] J. Mitschke, J. Georg, I. Scholz, C. M. Sharma, D. Dienst, J. Bantscheff, B. Voß, C. Steglich, A. Wilde, J. Vogel, and W. R. Hess, "An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC 6803," *Pnas*, vol. 108, no. 5, pp. 2124–2129, 2011.
- [135] L. S. Waters and G. Storz, "Regulatory RNAs in Bacteria," *Cell*, vol. 28136, no. 4, pp. 615–628, 2009.
- [136] W. C. Winkler and R. R. Breaker, "Genetic control by metabolite-binding riboswitches.," *ChemBioChem a Eur. J. Chem. Biol.*, vol. 4, no. 10, pp. 1024–1032, 2003.
- [137] B. J. Tucker and R. R. Breaker, "Riboswitches as versatile gene control elements," *Curr. Opin. Struct. Biol.*, vol. 15, no. 3, pp. 342–348, 2005.
- [138] R. R. Breaker, "Prospects for Riboswitch Discovery and Analysis," *Mol. Cell*, vol. 43, no. 6, pp. 867–879, 2011.
- [139] C. E. Lünse, A. Schüller, and G. Mayer, "The promise of riboswitches as potential antibacterial drug targets," *Int. J. Med. Microbiol.*, vol. 304, no. 1, pp. 79–92, 2014.
- [140] A. Ramesh, "Second messenger-sensing riboswitches in bacteria," *Semin. Cell Dev. Biol.*, vol. 47–48, pp. 3–8, 2015.
- [141] S. Gottesman and G. Storz, "Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations," *Cold Spring Harb. Perspect. Biol.*, vol. 3, 2011.
- [142] T. Morita, Y. Mochizuki, and H. Aiba, "Translational repression is sufficient for gene silencing by bacterial small noncoding RNAs in the absence of mRNA destruction.," *Pnas*, vol. 103, no. 13, pp. 4858–63, 2006.
- [143] C. M. Sharma, F. Darfeuille, T. H. Plantinga, and J. Vogel, "A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites," *Genes Dev.*, vol. 21, no. 21, pp. 2804–2817, 2007.
- [144] B. Vecerek, I. Moll, and U. Bläsi, "Control of Fur synthesis by the non-coding RNA RyhB and iron-responsive decoding," *EMBO J.*, vol. 26, no. 4, pp. 965–975, 2007.
- [145] K. Prévost, H. Salvail, G. Desnoyers, J. F. Jacques, É. Phaneuf, and E. Massé, "The small RNA RyhB activates the translation of *shiA* mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis," *Mol. Microbiol.*, vol. 64, no. 5, pp. 1260–1273, 2007.

- [146] J. H. Urban and J. Vogel, "Two Seemingly Homologous Noncoding RNAs Act Hierarchically to Activate *glmS* mRNA Translation," *PLoS Biol.*, vol. 6, no. 3, pp. 0631–0642, 2008.
- [147] N. Obana, Y. Shirahama, K. Abe, and K. Nakamura, "Stabilization of *Clostridium perfringens* collagenase mRNA by VR-RNA-dependent cleavage in 5' leader sequence," *Mol. Microbiol.*, vol. 77, no. 6, pp. 1416–1428, 2010.
- [148] E. Ramirez-Peña, J. Treviño, Z. Liu, N. Perez, and P. Sumby, "The group A *Streptococcus* small regulatory RNA FasX enhances streptokinase activity by increasing the stability of the *ska* mRNA transcript," *Mol. Microbiol.*, vol. 78, no. 6, pp. 1332–1347, 2010.
- [149] C. Lucchetti-Miganeh, E. Burrowes, C. Baysse, and G. Ermel, "The post-transcriptional regulator CsrA plays a central role in the adaptation of bacterial pathogens to different stages of infection in animal hosts," *Microbiology*, vol. 154, no. 1, pp. 16–29, 2008.
- [150] K. Lapouge, M. Schubert, F. H. T. Allain, and D. Haas, "Gac/Rsm signal transduction pathway of  $\gamma$ -proteobacteria: from RNA recognition to regulation of social behaviour," *Mol. Microbiol.*, vol. 67, no. 2, pp. 241–253, 2008.
- [151] E. Sonnleitner, T. Sorger-Domenigg, M. J. Madej, S. Findeiss, J. Hackermüller, A. Hüttenhofer, P. F. Stadler, U. Bläsi, and I. Moll, "Detection of small RNAs in *Pseudomonas aeruginosa* by RNomics and structure-based bioinformatic tools," *Microbiology*, vol. 154, no. 10, pp. 3175–3187, 2008.
- [152] L. Yuste, A. B. Hervás, I. Canosa, R. Tobes, J. I. Jiménez, J. Nogales, M. M. Pérez-Pérez, E. Santero, E. Díaz, J. L. Ramos, V. De Lorenzo, and F. Rojo, "Growth phase-dependent expression of the *Pseudomonas putida* KT2440 transcriptional machinery analysed with a genome-wide DNA microarray," *Environ. Microbiol.*, vol. 8, no. 1, pp. 165–177, 2006.
- [153] S. Frank, J. Klockgether, P. Hagendorf, R. Geffers, U. Schöck, T. Pohl, C. F. Davenport, and B. Tümmler, "*Pseudomonas putida* KT2440 genome update by cDNA sequencing and microarray transcriptomics," *Environ. Microbiol.*, vol. 13, no. 5, pp. 1309–1326, 2011.
- [154] J. Kim, J. C. Oliveros, P. I. Nikel, V. de Lorenzo, and R. Silva-Rocha, "Transcriptomic fingerprinting of *Pseudomonas putida* under alternative physiological regimes," *Environ. Microbiol. Rep.*, vol. 5, no. 6, pp. 883–891, 2013.
- [155] P. I. Nikel, J. Kim, and V. de Lorenzo, "Metabolic and regulatory rearrangements underlying glycerol metabolism in *Pseudomonas putida* KT2440," *Environ. Microbiol.*, vol. 16, no. 1, pp. 239–254, 2013.

## 5. Research Articles

This section contains the Research Articles included in this Ph.D. thesis.

### Research Article 1

**I. D'Arrigo**, K. Bojanovič, X. Yang, M. H. Rau, K. S. Long. (2016). Genome-wide mapping of transcription start sites yields novel insights into the primary transcriptome of *Pseudomonas putida*. *Environmental Microbiology*. Manuscript in press.

### Research Article 2

K. Bojanovič, **I. D'Arrigo**, K. S. Long. (2016). Global transcriptional responses to oxidative, osmotic, and imipenem stress conditions in *Pseudomonas putida*. Manuscript in preparation.

### Research Article 3

**I. D'Arrigo**, J. G. R. Cardoso, M. Rennig, N. Sonnenschein, M. J. Herrgård, K. S. Long. (2016). Investigation of *Pseudomonas putida* KT2440 transcriptome in different carbon sources reveals novelty in bacterial uptake systems. Manuscript in preparation.



## **Research Article 1**

**Genome-wide mapping of transcription start sites yields novel insights into the primary transcriptome of *Pseudomonas putida***

**I. D'Arrigo**, K. Bojanovič, X. Yang, M. H. Rau, K. S. Long. (2016). *Environmental Microbiology*. Manuscript in press.



**Genome-wide mapping of transcription start sites yields novel insights into the primary transcriptome of *Pseudomonas putida***

Isotta D'Arrigo<sup>1</sup>, Klara Bojanovič<sup>1</sup>, Xiaochen Yang<sup>1</sup>, Martin Holm Rau<sup>1</sup>, Katherine S. Long<sup>1,2</sup>

<sup>1</sup>Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Allé 6, DK-2970 Hørsholm, Denmark

<sup>2</sup>Address correspondence to: KSL [kalon@biosustain.dtu.dk](mailto:kalon@biosustain.dtu.dk) Phone: +45 45258024 Fax: +45 45258001

Keywords: *Pseudomonas putida*, KT2440, transcription start site, dRNA-seq, exonuclease, 5'UTR, riboswitch, TPP, sRNA

Running title: The primary transcriptome of *P. putida* KT2440



## Summary

The environmental bacterium *Pseudomonas putida* is an organism endowed with a versatile metabolism and stress tolerance traits that are desirable in an efficient production organism. In this work, differential RNA sequencing was used to investigate the primary transcriptome and RNA regulatory elements of *P. putida* strain KT2440. A total of 7937 putative transcription start sites (TSSs) were identified, where over two-thirds were located either on the opposite strand or internal to annotated genes. For TSSs associated with mRNAs, sequence analysis revealed a clear Shine-Dalgarno sequence but a lack of conserved overrepresented promoter motifs. These TSSs defined approximately fifty leaderless transcripts and an abundance of mRNAs with long leader regions of which eighteen contain RNA regulatory elements from the Rfam database. The thiamine pyrophosphate riboswitch upstream of the *thiC* gene was examined using an *in vivo* assay with GFP-fusion vectors and shown to function via a translational repression mechanism. Furthermore, fifty-six novel intergenic small RNAs and eight putative actuator transcripts were detected, as well as eight novel open reading frames (ORFs). This study illustrates how global mapping of TSSs can yield novel insights into the transcriptional features and RNA output of bacterial genomes.

## Introduction

*Pseudomonas putida* is a ubiquitous Gram-negative rod-shaped bacterium that has been used as a laboratory model for environmental bacteria and intensively studied regarding potential applications in industrial biotechnology (Poblete-Castro et al., 2012). Its simple nutritional requirements enable it to thrive in a wide variety of water and soil environments, including strains that colonize the rhizosphere and sites contaminated with chemical waste. Although features such as a versatile intrinsic metabolism, general robustness towards stress, and the ability to synthesize bioactive secondary metabolites are shared with other pseudomonads, *P. putida* is non-pathogenic and lacks the virulence factors harbored by other members of the genus that are human and plant pathogens (Nikel et al., 2014). A notable trait of *P. putida* is a superior tolerance to organic solvents (Ramos et al., 2015), as well as the ability of some strains to metabolize xenobiotic compounds. These characteristics combined with the availability of tools for genetic manipulation make *P. putida* an attractive host for heterologous gene expression and cell factory for the recombinant biosynthesis of natural products (Loeschcke and Thies, 2015).

One of the best characterized *P. putida* strains is KT2440 (Regenhardt et al., 2002), a plasmid-free derivative of the toluene-degrading strain mt-2 (Nakazawa, 2002). It is the preferred host for genetic manipulation and has been certified as a biosafety strain (Federal Register, 1982), a status that allows for industrial-scale production. The 6.2 Mb genome sequence confirmed the avirulence of the strain, and enabled a greater understanding of its physiology and metabolic repertoire (Nelson et al., 2002; Belda et al., 2016). Several genome-scale metabolic models have been developed and used to investigate the potential of the strain for the

production of biochemicals (Nogales et al., 2008; Puchalka et al., 2008; Sohn et al., 2010; Oberhardt et al., 2011; Belda et al., 2016).

RNA sequencing (RNA-seq) technology has emerged in recent years as the method of choice for transcriptome analysis and was used in an earlier study of *P. putida* KT2440 (Frank et al., 2011). The method of differential RNA sequencing (dRNA-seq) (Sharma and Vogel, 2014) distinguishes primary and processed transcripts, enabling the global determination of transcription start sites (TSSs). As these define 5' untranslated regions (5'UTRs) of mRNA transcripts, dRNA-seq facilitates the mapping and annotation of RNA regulatory elements in leader regions, including *cis*-acting metabolite-binding riboswitches. To date, over twenty families of riboswitches with known cognate ligands have been identified (Lunse et al., 2014; Ramesh, 2015), as well as several orphan riboswitches with unknown ligands (Breaker, 2011). Comparative genomics approaches have been used to reveal a number of conserved RNA motifs in bacteria and archaea. Although RNA elements have been also predicted in *Pseudomonas* spp. (Weinberg et al., 2007; Naville and Gautheret, 2010; Weinberg et al., 2010) and *P. putida* KT2440 (Frank et al., 2011; Sun et al., 2013), no studies with detailed characterization have been reported thus far.

In this work, a dRNA-seq approach was employed to gain insights into the *P. putida* KT2440 transcriptome, including *cis*-regulatory elements in 5'UTRs. Nearly 8000 TSSs were identified in four different growth conditions, where the majority were located either opposite of or internal to annotated genes. For the TSSs associated with mRNAs, sequence analysis showed a conserved Shine-Dalgarno sequence in leader regions but a lack of overrepresented sequence motifs in promoter regions. The study documents the discovery of roughly 50 leaderless mRNAs and hundreds of mRNAs with long leader regions, where the latter include

18 conserved RNA regulatory elements. Using plasmid reporter fusions, a TPP riboswitch element is demonstrated to function via a translational repression mechanism. This is, to our knowledge, the first *in vivo* riboswitch characterization in *P. putida*.

## **Results and discussion**

### *Experimental approach*

The dRNA-seq approach was used to map and investigate TSSs in *P. putida* strain KT2440. The method, described previously by Sharma and colleagues (Sharma et al., 2010), is based on the use of exonuclease enzyme that specifically degrades 5′monophosphorylated RNAs but not 5′triphosphorylated RNAs from the total RNA sample. This leads to a selective depletion of the processed RNAs and enrichment of primary RNA transcripts in the exonuclease-treated sample relative to the untreated total RNA sample. A comparison of sequencing reads between the treated and untreated samples allows for genome-wide TSS identification and improvement of the genome annotation.

Cells were grown in M9 minimal medium with either glucose or citrate as sole carbon sources and harvested in exponential and stationary phases (Fig. S1). Citrate was chosen as *P. putida* uses organic acids and amino acids, abundant in plant root exudates, as preferred carbon sources (Vilchez et al., 2000; Lugtenberg et al., 2001; Revelles et al., 2005), and the capability to grow on citrate is a feature distinct from *Escherichia coli*, which cannot normally use citrate as an energy source under oxic conditions (Blount et al., 2008). Unlike the model bacteria *E. coli* and *Bacillus subtilis*, glucose is not the favorite carbon source for pseudomonads (Rojo, 2010). In *P. putida*, succinate is consumed faster than glucose when the two substrates are provided simultaneously, although glucose is still well assimilated compared with other carbon

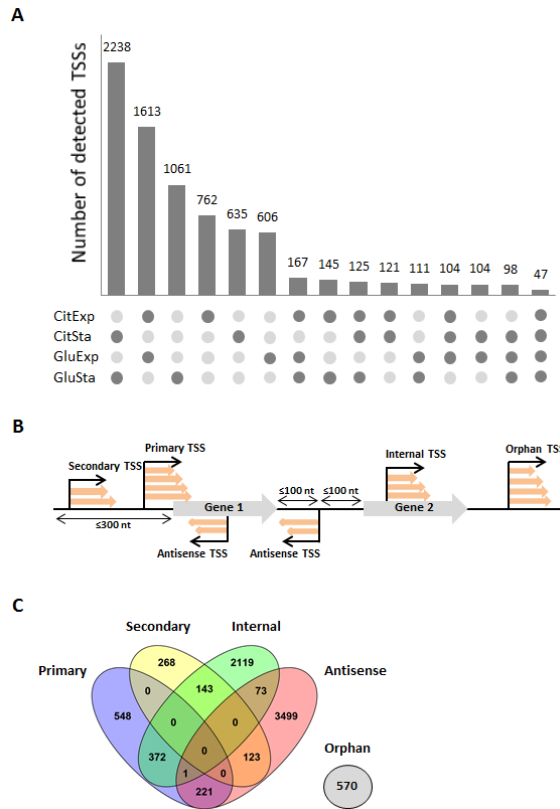
sources (La Rosa et al., 2015b; La Rosa et al., 2015a). Furthermore, the two chosen substrates, glucose and citrate, activate either glycolytic or gluconeogenic physiological regimes (Munoz-Elias and McKinney, 2006; Chavarria et al., 2013), respectively, leading to the expression of distinct metabolic genes (Kim et al., 2013).

Total RNA was extracted and split into two equal parts, where one was treated with exonuclease (Fig. S2). The untreated and treated RNA samples were used for preparation of strand-specific cDNA libraries for sequencing on the Illumina HiSeq platform. Strand-specific sequencing resulted in 2.5 to 14.6 million reads per sample, where on average 97% of the reads mapped to the reference chromosome (Table S1). Although the majority of reads mapped to rRNA, comparison of the percentage of rRNA reads between respective untreated and treated samples indicated a reduction in processed transcripts in the exonuclease-treated samples (Table S1). Moreover, despite the presence of ribosomal RNA, a sufficient number of reads were generated for transcriptional mapping in each condition (Creecy and Conway, 2015).

*Identification of transcription start sites: the majority are antisense or internal relative to annotated genes*

Transcription start sites were identified using TSSpredator software (Dugar et al., 2013), which normalizes the expression data and detects TSSs at genomic positions where a significant number of reads show major enrichment in the treated compared to the untreated samples. A total of 7937 putative TSSs were predicted in the *P. putida* KT2440 genome (Table S2), of which 762 (10%) were unique in exponential phase and 635 (8%) unique in stationary phase samples for cells grown in citrate (Fig. 1A). For cells grown in the presence of glucose, 606 (8%) and 1061 (13%) unique putative TSSs were predicted in exponential and

stationary phase samples, respectively. Nearly half of the TSSs (48%) were identified in the same growth phase, including 2238 TSSs (28%) in stationary phase and 1613 (20%) in exponential phase for both carbon sources. However, 42% of TSSs were clearly associated with growth in either glucose or citrate as sole carbon sources. Surprisingly, only 47 TSS (0.6%) were detected in all four conditions (Fig. 1A).



**Figure. 1.** Genome-wide identification of transcription start sites. A. The numbers of identified transcription start sites (TSSs) in the four conditions with either glucose (Glu) or citrate (Cit) as sole carbon sources in exponential (Exp) or stationary (Sta) growth phases are shown. Dark and light dots indicate the presence or absence of TSSs in a growth condition, respectively. B. Schematic illustration of categories used for TSS classification (including primary, secondary, internal, antisense and orphan TSS groups) based on their genomic context relative to annotated genes. C. Venn diagram showing the distribution of identified TSSs into the categories depicted in panel B. A TSS can be associated with more than one group.

These data show a similarity in the positions of transcription initiation between the two carbon sources when the same growth phase is considered. The biggest difference was observed between the exponential and stationary phase growth conditions, underscoring the vast changes in gene expression to respond to the different physiological status of the cell in the two growth phases. Therefore, the two growth phases introduce more variability than the two carbon sources. Additionally, concerning the TSSs found in only one condition, approximately 7% of the 3064 unique TSSs corresponded to TSSs with positions varying by ten or less nucleotides (nt) in different conditions. The rest were composed of TSSs corresponding to specific expression in one condition, or TSSs with major coordinate differences in different conditions.

The identified TSSs were classified based on their putative origin and genomic context into primary, secondary, internal, antisense and/or orphan TSSs categories (Fig. 1B) (Dugar et al., 2013). TSSs located within 300 nt upstream of an annotated gene were designated as primary and secondary, with the former exhibiting the most cDNAs and strongest expression relative to the latter. TSSs located within and on the same strand of annotated genes were defined as internal, whereas antisense TSSs were positioned either inside or at a maximal distance of 100 nt relative to annotated genes on the opposite strand. Orphan TSSs were not in close proximity to annotated genes and belonged to none of the aforementioned categories (Fig. 1B and 1C). Surprisingly, the numbers of primary and secondary TSSs were considerably lower than the numbers of internal and antisense TSSs. One factor contributing to the low number of primary TSSs is that these could be placed in other TSS categories such as internal and orphan in the case of long leaders with lengths greater than 300 nt. This suggests that a threshold length of 300



nt upstream of start codons for definition of primary and secondary TSSs, although sufficient in the organisms studied previously (Irnov et al., 2010; Sharma et al., 2010; Filiatrault et al., 2011; Kroger et al., 2012; Sahr et al., 2012; Schmidtke et al., 2012; Wurtzel et al., 2012; Dugar et al., 2013; Wiegand et al., 2013; Kopf and Hess, 2015; Nuss et al., 2015; Papenfort et al., 2015), is not optimal for all bacteria and that longer leader regions may be more prevalent in *P. putida*.

Nearly half of the identified TSSs were classified as antisense, indicating a high level of transcription initiation on the antisense strand. Antisense transcription is now highly reported in transcriptome analysis, and several studies have revealed the presence of a large number of transcripts antisense to annotated genes and the 5' or 3' ends of mRNAs in different organisms (Georg et al., 2009; Liu et al., 2009; Toledo-Arana et al., 2009; Dornenburg et al., 2010; Filiatrault et al., 2010). Acting via extensive base pairing, antisense RNAs regulate the expression of the gene on the opposite strand by modulating the transcription, stability or translation of the specific target. In some cases, antisense RNAs can play a dual role by also functioning as an mRNA encoding a small protein (Silby and Levy, 2008), or regulating several genes other than the target on the opposite strand with the RNA chaperone Hfq (Opdyke et al., 2004; Mandin et al., 2007; Arnvig and Young, 2009). Despite the high number of antisense transcripts identified, only a few have been functionally characterized. Besides antisense transcripts acting as RNA regulators, antisense transcription may also represent noise due to nonspecific transcription in bacteria (Georg and Hess, 2011; Thomason et al., 2015). At least one antisense TSS was identified for 1991 (36%) of *P. putida* KT2440 genes. Previous studies identified putative antisense RNAs for 12% of all genes in *Mycoplasma pneumoniae* (Guell et al., 2009), 2% in *Sinorhizobium meliloti* (Schluter et al., 2010), less than 1%

in *Bacillus subtilis* (Irnov et al., 2010), and 22% in *Escherichia coli* (Dornenburg et al., 2010), without any further analysis to validate the presence of the antisense RNAs. Similar to *P. putida*, *Helicobacter pylori* was reported to have antisense RNAs on 46% of all genes, of which 21 RNA transcripts were supported with additional experiments (Sharma et al., 2010).

A direct comparison with other TSS identification studies is not straightforward due to different experimental approaches and TSS classification schemes. However, three previous studies used the same experimental approach (dRNA-seq methodology followed by Illumina sequencing and TSSpredator analysis) in different bacteria and growth conditions (Dugar et al., 2013; Bischler et al., 2015; Thomason et al., 2015). *P. putida* has the highest percentage of orphan TSSs (6%) compared to *Escherichia coli* (3%), *Campylobacter jejuni* (2%) and *Helicobacter pylori* (2%), suggesting the presence of a large number of sRNA candidates and unannotated ORFs (Fig. S3). Even though a direct comparison with other TSS identification studies is not possible, it is worth mentioning that similar to *P. putida*, high percentages of orphan TSSs were previously observed in cyanobacteria (Kopf and Hess, 2015). Transcriptomic analysis of seven cyanobacteria by dRNA-seq revealed high levels of transcription in non-coding regions among all the TSSs identified, where the percentage of orphan TSSs varied between 5.1% and 26.7% depending on the organism (Mitschke et al., 2011; Voss et al., 2013; Kopf et al., 2014; Pfreundt et al., 2014; Voigt et al., 2014; Kopf et al., 2015).

In order to confirm the TSS predictions, the full-length sequences of three selected gene transcripts were determined by 5'RACE (rapid amplification of cDNA ends). The TSS predictions were validated in the following genes with high expression levels or differential expression

between growth on glucose and citrate: PP0147, a citrate transporter; PP4010, cold-shock protein D (*cspD*); and PP1623, the RNA polymerase sigma factor (*rpoS*). In all cases there was good agreement between RACE determined and predicted TSS positions, with a maximum divergence of 9 nucleotides (Table S3). This lends reliability to the TSS predictions and the potential of TSSpredator as a valid automated TSS prediction tool. Nevertheless, the accuracy of the TSSs coordinates can be improved by modifying the library preparation protocol in order to reduce the relative amount of rRNA and by increasing the sequencing depth. Another possibility is to increase sensitivity in TSSpredator parameters, leading to a higher number of identified TSSs but also an increase in false positives. Therefore, the TSSpredator parameters (see Experimental Procedures) chosen for this study were an optimal compromise between the number and accuracy of TSSs coordinates identified. Moreover, the 5'RACE result for the *rpoS* transcript revealed a primary TSS located 369 nucleotides upstream of the ATG start codon (Fig. S4). This TSS was positioned inside the upstream PP1622 gene and therefore classified as an internal TSS. Thus, not all primary TSSs are within 300 nucleotides of the start codon and some internal TSSs may function as the primary TSS of the downstream mRNA. This finding can account in part for the high number of internal TSSs relative to primary and secondary TSSs identified in *P. putida*.

A common feature of bacterial genomes is the polycistronic organization of several genes in an operon, where sets of co-regulated and co-transcribed genes are transcribed as a single mRNA, allowing rapid adaptation to environmental changes (Lawrence, 2002). A total of 1076 multi-gene operons were predicted here in *P. putida* (Table S4) that were mostly composed of two (59%) or three (20%) genes and 7 operons included more than 10 genes (Fig. S5). Considering that *P. putida*

KT2440 has 5350 coding sequences and 3120 (58%) were predicted to be organized in multi-gene operons, the remaining 2230 (42%) could be single-gene operons or not expressed under the studied conditions and therefore not categorized. Although information on the positions of 3' ends are also required for a precise mapping of operons, the above estimation is based on the pattern of expressed genes under the studied conditions and is similar to the single- and multi-gene operon composition of *E. coli* (Conway et al., 2014).

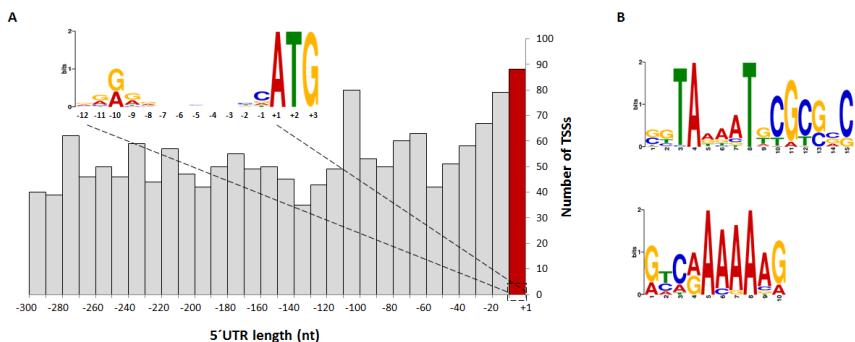
Therefore, there are several possible explanations for the absence of identified primary TSSs for some annotated genes and the relative low number of TSSs in this class including: a) the gene is not expressed under the tested conditions, b) the TSS is longer than 300 nt and thus it is not classified as a primary TSS but either an orphan or internal TSS, c) the gene is part of an operon so it is co-transcribed and shares the TSS with the upstream gene and d) the gene is part of an operon with an internal promoter and terminator and therefore the TSS is classified as internal.

#### *Investigation of 5' untranslated regions reveals an abundance of mRNAs with long leaders*

A total of 1676 primary and secondary TSSs were identified, of which 1599 were associated with mRNAs defining the 5'UTR regions of protein-coding genes, and 77 were related to RNAs transcripts (rRNA and tRNA). The 5'UTRs or leader regions of mRNA, defined by the transcription start site and the nucleotide just before the start codon were examined in *P. putida*. Leaderless mRNAs in bacteria with an mRNA starting at the first codon or up to ten nucleotides upstream were once considered rare (Moll et al., 2002; Laursen et al., 2005), but recent studies have shown that they are much more common (Brock et al.,

2008; Sharma et al., 2010; Sahr et al., 2012; Schmidtke et al., 2012; Cortes et al., 2013; Schluter et al., 2013; Kopf et al., 2015; Shell et al., 2015). Out of the 1599 TSSs associated with mRNAs, 88 defined mRNAs with a 5'UTR of ten nucleotides or less. Of these, 51 mRNAs were leaderless with no other longer TSS identified (Table S5) and the other 37 mRNAs had both leadered and leaderless variants. A significant fraction of the leaderless mRNAs encode gene products with functions related to nucleic acids (DNA binding proteins, DNA/RNA modification and nucleotide synthesis enzymes) or of unknown function. This result indicates that leaderless transcripts may be more frequent in certain gene function categories than others; for instance information storage and processing categories tend to have a higher fraction of leaderless mRNAs than metabolic-related genes (Nakagawa et al., 2010; Zheng et al., 2011). A previous study on the *P. putida* KT2440 transcriptome identified eight leaderless mRNAs out of 170 highly expressed 5'UTRs (Frank et al., 2011). In this study leader regions longer than 100 nucleotides are reported for six of these transcripts. The larger number of leaderless mRNAs identified here is likely due to the different conditions examined and the dRNA-seq approach that specifically reveals features of 5'UTRs (Sharma and Vogel, 2014).

The 5'UTR length distribution of the TSSs associated with mRNA is shown in Fig. 2A.



**Figure. 2.** Leader regions and analysis of promoter motifs. A. Plot showing the distribution of 5'UTR lengths based on 1599 primary and secondary TSS of mRNAs. 5'UTRs with lengths of 10 nt or less are shown in red. The insert shows the overrepresented motif for the 5'UTRs, and consists of the Shine-Dalgarno sequence and the start codon. B. Two overrepresented motifs found in promoter regions, including the -10 box (top) and the A<sub>5</sub> sequence (bottom).

There is a median 5'UTR length of 136 nt and a high number of 5'UTRs with lengths between 100 - 300 nt. Nearly 1000 leaders are longer than 100 nt, and about half of these are longer than 200 nt. This is an underestimate as there are likely to be a significant number of mRNAs with leaders longer than 300 nucleotides that are not taken into account here. This differs from previously reported distributions in other organisms, which have median 5'UTR lengths between 33 - 54 nt and the highest 5'UTR frequencies in the 20 - 40 nt range using a variety of methods (Irnov et al., 2010; Sharma et al., 2010; Filiatrault et al., 2011; Kroger et al., 2012; Sahr et al., 2012; Schmidtke et al., 2012; Wurtzel et al., 2012; Dugar et al., 2013; Wiegand et al., 2013; Kopf and Hess, 2015; Nuss et al., 2015; Papenfort et al., 2015). Our result shows that *P. putida* KT2440 has many potential mRNAs with long 5'UTRs compared with

other bacteria examined to date. These long leader regions may mediate regulation on downstream genes via specific RNA secondary structures (Winkler and Breaker, 2003; Araujo et al., 2012), such as *cis*-acting riboswitches (Coppins et al., 2007) and be targeted via base-pairing interactions with *trans*-acting sRNA regulators (Waters and Storz, 2009). Therefore, extended 5'UTR regions in *P. putida* suggest a high potential for mRNA regulation and the presence of *cis*-regulatory elements.

The sequences adjacent to the TSSs of mRNAs were investigated for the presence of Shine-Dalgarno sequences by using the Multiple EM for Motif Elicitation (MEME) tool for motif discovery (Bailey et al., 2009). The Shine-Dalgarno sequence was searched for in the regions surrounding the start codons using genomic sequences corresponding to 40 nucleotides upstream and downstream of the start codon as input. The Shine-Dalgarno sequence was clearly identified within the 12 nucleotides upstream of the ATG start codon (Fig. 2A).

The promoter regions were also investigated by MEME using the sequences 50 nucleotides upstream of the transcription start site. For this search, TSSs of tRNA and rRNA genes were also included (1676 total input sequences). This yielded two motifs with significant E-values ( $< 10^{-30}$ ) including a possible -10 box and an A<sub>5</sub> sequence (Fig. 2B). Interestingly, the motifs were found in surprisingly few input sequences, 77 and 63, for the -10 box and A<sub>5</sub> motifs, respectively. Moreover, the motifs did not have a specific nucleotide position, as their locations varied relative to the TSS between the different sequences. A previous transcriptomic study found the pentameric polyA motif but neither a -10 nor the -35 region motifs (Frank et al., 2011). These results emphasize the relative lack of overrepresented promoter motifs with a clear position for *P. putida* transcripts. It has been noted previously that TSS neighborhoods can be highly heterogeneous with different promoter

architectures affecting the position of transcription initiation depending on the growth phase and the environment (Narlikar, 2014). Therefore the difficulty in finding conserved overrepresented motifs in specific locations in *P. putida* promoter regions suggests the presence of a different promoter architecture and high level of heterogeneity surrounding the TSS.

#### *Cis-regulatory RNA elements in 5'UTR regions*

As 5'UTRs contain elements that exert *cis*-regulation on downstream genes, the identification and characterization of these elements can contribute to the understanding of bacterial adaptation under different conditions. The high number of mRNAs with long leader regions in *P. putida* KT2440 prompted the investigation of these 5'UTRs for possible regulatory functions. In this respect much is still unknown for *P. putida*, where the *cis*-regulatory elements known to date have been based on sequence comparison of transcriptomic data (Frank et al., 2011) and comparative genomic analysis (Weinberg et al., 2007; Weinberg et al., 2010; Sun et al., 2013), but further characterization is lacking.

Therefore, 5'UTRs defined by primary and secondary TSSs with lengths of 80-300 nucleotides were investigated for possible *cis*-regulatory RNA structures by searching for homologies with annotated sequences in Rfam databases (Nawrocki et al., 2015). The input sequences included those from 100 nucleotides upstream of the identified TSS to 50 nucleotides downstream of the first codon. In addition to the *in silico* search, manual inspection of TSS read profiles was performed with Integrative Genomics Viewer (IGV) (Robinson et al., 2011; Thorvaldsdottir et al., 2013). A total of eighteen *cis*-RNA regulatory elements were found with homology to known RNA motifs using the Rfam database (Table 1).



**Table 1.** *Cis*-RNA regulatory elements in 5'UTRs.

Number	Rfam motif	Predicted TSS position <sup>a</sup>	Strand	Downstream gene number and annotation	Reference
1	<i>gabT</i>	85	+	PP0214 <i>gabT</i> : 4-aminobutyrate aminotransferase	Weinberg <i>et al.</i> 2010 ( <i>Pseudomonas</i> )
2	<i>rpsL</i> pseudo	95	+	PP0449 <i>rpsL</i> : 30S ribosomal protein S12	Naville <i>et al.</i> 2010 ( <i>Pseudomonadaceae</i> )
3	Alpha RBS	238	+	PP0476 <i>rpsM</i> : 30S ribosomal protein S13	Schlax <i>et al.</i> 2001 ( <i>E. coli</i> )
4	FMN	186; 196	-	PP0530 <i>ribB</i> : 3,4-dihydroxy-2-butanone 4-phosphate synthase	Frank <i>et al.</i> 2011 ( <i>P. putida</i> KT2440)
5	YybP-YkoY leader*	178	-	PP0760 hypothetical protein	Frank <i>et al.</i> 2011 ( <i>P. putida</i> KT2440)
6	2 group II (D1D4-3 and D1D4-1)**	279	+	PP1250 group II intron-encoding maturase	Lehmann & Schmidt 2003
7	Pseudomon- <i>groES</i> RNA	111	+	PP1360 <i>groES</i> : co-chaperonin <i>GroES</i>	Weinberg <i>et al.</i> 2010 ( <i>Pseudomonas</i> )
8	Cobalamin*	246	+	PP1672 <i>cobO</i> : cob(I)yrinic acid a,c-diamide adenosyltransferase	Frank <i>et al.</i> 2011 ( <i>P. putida</i> KT2440)
9	<i>gyrA</i> RNA	122; 148	+	PP1767 <i>gyrA</i> : DNA gyrase subunit A	Weinberg <i>et al.</i> 2010 ( <i>Pseudomonas</i> )
10	2 group II (D1D4-3 and D1D4-1)**	280	+	PP1846 group II intron-encoding maturase	Lehmann & Schmidt 2003

11	Cobalamin*	197	-	PP2418	hypothetical protein	Sun <i>et al.</i> 2013 ( <i>P. putida</i> KT2440)
12	TPP	135	+	PP3185	<i>pet18</i> : TenA family transcriptional activator	Sun <i>et al.</i> 2013 ( <i>P. putida</i> KT2440)
13	Cobalamin	217; 248	-	PP3508	<i>cobW</i> : cobalamin biosynthesis protein <i>CobW</i>	Sun <i>et al.</i> 2013 ( <i>P. putida</i> KT2440)
14	<i>sucA-II</i> RNA	110; 235	-	PP4189	<i>sucA</i> : 2-oxoglutarate dehydrogenase E1	Weinberg <i>et al.</i> 2010 (Pseudomonadales)
15	Ribosomal S15 leader*	107	-	PP4709	30S ribosomal protein S15	Sun <i>et al.</i> 2013 ( <i>P. putida</i> KT2440)
16	TPP	246; 252	-	PP4922	<i>thiC</i> : thiamine biosynthesis protein <i>ThiC</i>	Frank <i>et al.</i> 2011 ( <i>P. putida</i> KT2440)
17	SAH	152	+	PP4976	<i>ahcY</i>	Weinberg <i>et al.</i> 2007 (Proteobacteria)
18	Pseudomon-Rho	128; 136	-	PP5214	<i>rho</i> : transcription termination factor <i>Rho</i>	Weinberg <i>et al.</i> 2010 ( <i>Pseudomonas</i> )

---

<sup>a</sup> The TSS position is reported as the leader length or number of nucleotides upstream of the downstream gene.

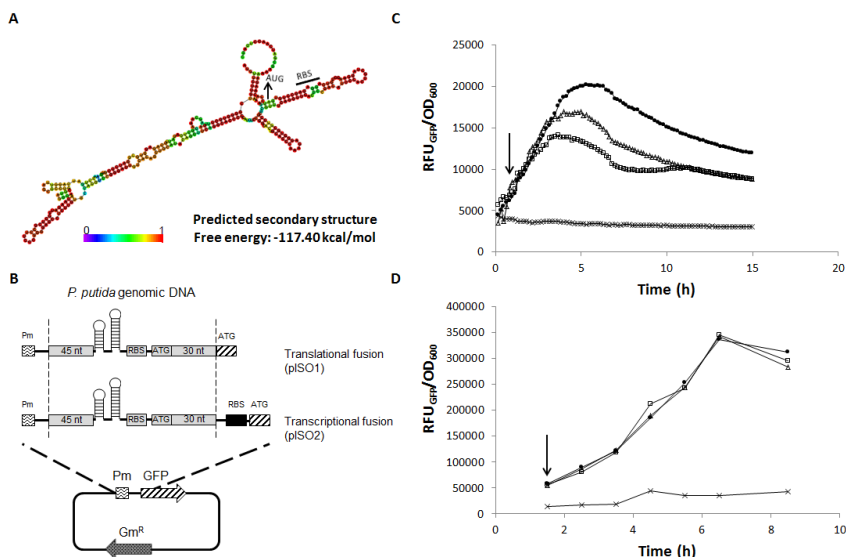
\* Regulatory element found by visual inspection on IGV profiles.

\*\* Ribozyme.

The riboswitch elements predicted previously in *P. putida* KT2440 by transcriptomics (Frank et al., 2011) and comparative genomics (Sun et al., 2013) were confirmed here. Moreover, other relevant motifs were also found, including those related to the genus *Pseudomonas* (Weinberg et al., 2010) and phylum Proteobacteria (Weinberg et al., 2007; Naville and Gautheret, 2010), and two group II catalytic RNA (ribozymes) elements that occur in all domains of life (Lehmann and Schmidt, 2003). Finally, an RNA element associated with repression of the ribosomal protein operon S13 by the ribosomal protein S4 was identified in *P. putida* KT2440 that was previously described in *E. coli* (Schlax et al., 2001).

*The TPP riboswitch upstream of the thiC gene acts via a translational inhibition mechanism*

The TPP riboswitch upstream of the *thiC* gene was chosen from the list of predicted *cis*-regulatory RNA elements (Table 1) as it has been characterized in other bacteria and suitable for further investigation of the ligand-dependent regulatory mechanism in *P. putida*. The identified TPP sequence folded in a stem-loop structure stabilized by a negative free energy value ( $\Delta G = -117.40$  kcal/mol), predicted by RNAfold WebServer (Gruber et al., 2008) (Fig. 3A).



**Figure. 3.** Characterization of the TPP riboswitch. The predicted TPP riboswitch upstream of the *thiC* gene is characterized for its ligand-dependent regulatory mechanism by GFP reporter fusion systems. **A.** The secondary structure of the TPP riboswitch predicted by RNAfold WebServer (Gruber et al., 2008). The base coloring represents the base-pairing probability. The free energy of the conformation is reported. **B.** Representation of the two reporter fusion systems: the translational (pISO1) and the transcriptional (pISO2) fusion plasmids. **C, D.** Fluorescence levels in the absence and presence of the ligands (TPP and thiamine) in the translational fusion (**C**) and transcriptional fusion (**D**) plasmids. There is a repression of fluorescence in the translational fusion when either TPP or thiamine are added. However, fluorescence levels in the transcriptional fusion are unchanged by ligand addition. Arrows indicate the points of ligand addition. RFU/OD<sub>600</sub> graphs are showed: curves with no induction and no ligand (×), curves with induction but absence of ligand (●), and curves with 0.5 mM TPP (Δ) or 5 μM thiamine (□) after induction.

The TPP riboswitch (or Thi-box) binds directly to its natural ligand TPP, the active form of thiamine (vitamin B1), and represses the expression of thiamine-related genes (Miranda-Rios, 2007). The regulatory mechanisms of the TPP riboswitches upstream of the *thiC* and *thiM* genes in *E. coli* (Winkler et al., 2002; Ontiveros-Palacios et al., 2008; Caron et al., 2012), and the *thi* operon (*tenA*) in *B. subtilis* (Mironov et al., 2002) have been studied in the presence of TPP and thiamine. The involved Thi-box elements in *E. coli* and *B. subtilis* differ in their mechanism of action, where they function via translational repression (Winkler et al., 2002) and transcription termination (Mironov et al., 2002) mechanisms, respectively. This has led to the suggestion that the TPP riboswitch induces transcription termination in Gram-positive bacteria and inhibits translation initiation in Gram-negative bacteria (Nudler and Mironov, 2004). Thus, it was of interest to test this hypothesis by investigating the regulatory function of the specific TPP riboswitch sequence predicted in *P. putida*.

The regulatory mechanism of the TPP riboswitch upstream of the *thiC* gene was tested by using a translational fusion with a GFP reporter in a plasmid construct. The resulting plasmid (pISO1) contained the inducible promoter Pm (Marques et al., 1998; Miura et al., 1998; Winther-Larsen et al., 2000), the natural genome sequence of *P. putida* KT2440 (including the putative riboswitch domain, the natural ribosome binding site (RBS) and 30 nt of the natural downstream gene *thiC*), and the GFP gene (Fig. 3B). The KT2440 (pISO1) strain was grown in a microtiter plate, induced with 3-methylbenzoate for GFP expression, followed by monitoring the level of fluorescence with and without TPP and thiamine ligands. The plasmid-transformed strain showed a reduction of relative fluorescence units (RFU) when TPP or thiamine were added to the media compared to the absence of ligand (Fig. 3C),

while the RFU levels in the plasmid-free strain were not affected. This confirmed the regulatory mechanism of the riboswitch sequence, which repressed the expression of the downstream GFP gene in the presence of either ligand. Different ligand concentrations between 10 nM and 1.5 mM were tested, and addition of 0.5 mM TPP and 5  $\mu$ M thiamine led to RFU reductions of 21% and 35% for TPP and thiamine, respectively. These ligand concentrations yielded the maximum extent of repression, as larger effects were not observed with higher ligand concentrations.

*In vitro* studies have demonstrated a stronger binding of the TPP ligand to the *thiC* Thi-box structure compared with the precursor thiamine, with the riboswitch exhibiting more than 1000-fold discrimination between the two ligands (Winkler et al., 2002; Yamauchi et al., 2005; Edwards and Ferre-D'Amare, 2006; Lang et al., 2007; Ontiveros-Palacios et al., 2008; Haller et al., 2013). Therefore the 100-fold lower thiamine concentration relative to TPP concentration required for repression of GFP expression observed in this study may be a consequence of differences in ligand uptake into the cell. Thiamine is synthesized by most prokaryotes (Begley et al., 1999; Jurgenson et al., 2009), and can alternatively be taken up from the environment (Webb et al., 1998), but the responsible transporters and cellular uptake mechanisms remain unclear in many organisms, including *P. putida* (Webb et al., 1998; Jurgenson et al., 2009; Rodionov et al., 2009; Jeanguenin et al., 2012; Rodionova et al., 2015). Therefore, the difference in the active concentrations of TPP and thiamine could be related to the specificity of the transporter. Specifically, the transporter may have higher affinity for thiamine that facilitates its entry, while TPP transport may be less efficient and require higher concentrations for activity.

To better understand the regulatory mechanism and confirm the expected translational inhibition of the TPP riboswitch, a plasmid vector

was constructed with a transcriptional fusion of the TPP motif and the GFP gene (pISO2). For this fusion, in addition to the natural RBS introduced with the riboswitch sequence from the *P. putida* genome, a second RBS was introduced just upstream of the reporter gene (Fig. 3B). In this construct, the translation repression activity of the riboswitch should only sequester the natural RBS but not the second RBS, and thus allow GFP expression in the presence or absence of ligand.

The RFU levels of *P. putida* KT2440 (pISO2) with and without ligands, were monitored during growth, and dilution factors were applied to avoid overflow measurement of fluorescence due to the two RBS sequences. The transcriptional fusion plasmid allowed a continuous expression of GFP and no repression of fluorescence was observed upon ligand addition (Fig. 3D). In this model, the reporter expression was dependent on the level of the mRNA and its translation regulated by the second RBS, which was not sequestered by the riboswitch structure. This confirms the translational repressor activity of the TPP riboswitch upstream of the *thiC* gene in *P. putida* KT2440. In the case of a mechanism involving transcription termination in the presence of the ligands, there would be no transcribed mRNA and consequently no expression of GFP. Our data supporting the translational repression mechanism of the TPP riboswitch in *P. putida* KT2440 is consistent with the hypothesis that in Gram-negative bacteria, Thi-box elements act by interfering with RBS-ribosome recognition instead of transcription termination. This work represents the first *in vivo* riboswitch characterization in *P. putida*.

*Identification of small RNA candidates derived from intergenic regions and 5'UTRs*

In addition to *cis*-regulatory RNA elements, eighty putative intergenic small RNA transcripts were identified based on computational prediction and visual inspection of expression profiles (Table S6) (Fig. S6B). Twenty-four transcripts were annotated previously or found to have homology to known sRNAs or RNA motifs in the Rfam database (Nawrocki et al., 2015). The other fifty-six sRNAs (named Pit for *P. putida* intergenic transcript) were novel sRNAs candidates, and a subset of a complete list of putative sRNAs identified in another study with a deeper sequencing depth (Bojanovič et al., manuscript in preparation). Moreover, three additional sRNAs (RNA1, RNA2, RNA3) were detected here but not in Bojanovič et al. (manuscript in preparation), likely due to the different library preparation strategies and the dRNA-seq approach used here. A previous study identified thirty-six sRNAs, of which fourteen were novel (Frank et al., 2011). In this study, six of the latter fourteen were detected (named as IGR in Table S6). The reason the other eight were not detected may be attributed to the expression of sRNAs only in specific growth conditions and differences in experimental protocols for RNA isolation and library construction as documented previously (Gomez-Lozano et al., 2012).

Further analysis with IGV revealed an additional eight transcripts (Table S7) with read profiles consistent with actuatons (Fig. S6C). Actuatons are a class of sRNAs characterized by a high number of reads in the 5'UTR and the presence of a terminator in the proximity of the downstream gene. The downstream mRNA lacks its own TSS and originates from terminator read-through, such that these sRNAs and their downstream mRNAs are joined in a unique transcriptional unit (Kopf



and Hess, 2015). This group of sRNAs is expected to function as possible regulators.

### *Identification of novel ORFs*

The TSS prediction revealed 570 orphan TSSs, which were used for the identification of putative novel ORFs in *P. putida* KT2440. The DNA sequences between the predicted orphan TSSs and the downstream annotated genes were collected and analyzed by the *in silico* gene finders GLIMMER (Salzberg et al., 1998; Delcher et al., 2007) and GeneMark (Lukashin and Borodovsky, 1998; Besemer and Borodovsky, 2005). The RNA-seq data were then used to confirm the transcription of the ORFs predicted by both GLIMMER and GeneMark. Twenty-one putative ORFs were identified (Table S8) and classified into two categories. In the first, the sequences from the two gene finders were completely overlapping, having the same translational start and stop positions. In the second, the sequences had different start sites predicted by GLIMMER and GeneMark but the same stop position. Of the twenty-one predicted ORFs, twelve belonged to the first and nine to the second category. The functions of the putative ORFs were evaluated by sequence homology in protein Blast (Johnson et al., 2008). Five ORFs were homologous to functional proteins or specific domains in *Pseudomonas* and other organisms (PP3108.2 and PP3108.4: rhs family protein, PP1810.1: DUF 3077 superfamily, PP1935.4: resolvase, PP2509.1: diadenosine tetraphosphate hydrolase), while the sixteen remaining ORFs were hypothetical proteins (Table S8).

From the twenty-one putative ORFs identified here, eight were also detected previously in Frank et al. (Frank et al., 2011) with exactly the same coordinates, and five with a different start position. The remaining eight are novel ORFs that have not been described previously (Table

S8). The different numbers of putative ORFs identified in the two studies can be due in part to different patterns of gene expression in the investigated conditions and also to the fact that the analysis performed here is limited to the orphan TSSs.

### **Concluding remarks**

This study is the first genome-wide TSS analysis in *P. putida* under different growth conditions, and provides a deeper understanding of its metabolic versatility and ability to adapt to different environments. The novel genomic features uncovered here prompt intriguing questions regarding promoter selection and variability in *P. putida* under different conditions, as well as the number of antisense transcripts and whether these are located opposite genes in specific functional categories. The hundreds of mRNAs with long leader regions highlight the issue of their biological function and the wider role of *cis*-regulation in *P. putida*. The structures of these leaders would be interesting to study with next-generation sequencing approaches to probe RNA structure on a global scale or the RNA structurome with the aim of identifying RNA thermometers (Righetti and Narberhaus, 2014) and other *cis*-regulatory elements such as riboswitches. The work underscores the complexity and diversity of bacterial transcriptomes and as the interest in *P. putida* as biotechnological tool is increasing, the genomic features identified here are a benchmark for future studies of its gene expression and metabolic engineering.

### **Experimental Procedures**

#### *Bacterial strains and growth conditions*

*Pseudomonas putida* strain KT2440 was used in all experiments. *P. putida* was cultivated in M9 minimal medium (per liter:

Na<sub>2</sub>HPO<sub>4</sub>·12H<sub>2</sub>O, 70 g; KH<sub>2</sub>PO<sub>4</sub>, 30 g; NH<sub>4</sub>Cl, 10 g; NaCl, 5 g) supplemented with ammonium iron citrate, magnesium sulfate, and trace metals (per liter: H<sub>3</sub>BO<sub>3</sub>, 300 mg; ZnCl<sub>2</sub>, 50 mg; MnCl<sub>2</sub>·4H<sub>2</sub>O, 30 mg; CoCl<sub>2</sub>, 200 mg; CuCl<sub>2</sub>·2H<sub>2</sub>O, 10 mg; NiCl<sub>2</sub>·6H<sub>2</sub>O, 20 mg; and NaMoO<sub>4</sub>·2H<sub>2</sub>O, 30 mg) (Abril et al., 1989). The medium included either sodium citrate (10 mM) or glucose (0.5% (w/v)) as sole carbon sources. All liquid cultures were grown at 30°C with vigorous shaking at 250 rpm. The cultures used for RNA isolation were grown from single colonies isolated from LB agar plates containing 25 µg/mL chloramphenicol grown overnight at 30°C. These were used to inoculate 5 mL M9 medium supplemented with chloramphenicol (25 µg/mL). The overnight cultures were diluted to a starting OD<sub>600</sub> of 0.05 in 100 mL M9 medium in 250 mL Erlenmeyer flasks.

*Escherichia coli* strain NEB5α (New England Biolabs) was used for cloning and propagation of plasmids (Table S9). Chemically competent cells of NEB5α were prepared as described elsewhere (Inoue et al., 1990) and had an estimated transformation efficiency of  $2.6\text{--}3.3 \times 10^7$  CFU/µg DNA. *E. coli* was propagated at 37°C in LB supplemented with gentamicin (10 µg/mL) when required.

#### *Cell harvest and RNA isolation*

*P. putida* KT2440 cells were harvested in mid-exponential phase (OD<sub>600</sub> ~ 0.5 and 1 for citrate and glucose cultures, respectively) and in early-stationary phase (OD<sub>600</sub> ~ 1.5 and 4.9 for citrate and glucose cultures, respectively). Cells were harvested by transferring 20 mL of each culture into 50-mL Falcon tubes containing 4 mL of stop solution (5% phenol in 95% ethanol, 4°C), vortexed for 15 sec and kept on ice for 5 min. Following centrifugation (8000 rpm, 2 min, 4°C) in a Multifuge X3 Fr centrifuge (Thermo Scientific), cells were resuspended in 2 mL of

supernatant by pipetting and split into two RNase-free 1.5 mL tubes. After centrifugation (7000 x g, 5 min, 4°C) and removal of supernatant, the pellet was dissolved in 1 mL of TRIzol Reagent (Invitrogen, Life Technologies), incubated 5 min at room temperature and stored at -80°C. Total RNA extraction and DNA removal by treatment with DNaseI were performed as previously described (Gomez-Lozano et al., 2012). The integrity of total RNA, the presence of rRNAs and tRNAs, as well as DNA contamination were assessed with a RNA 6000 Nano chip on Agilent 2100 Bioanalyzer (Agilent Technologies). Total RNA was extracted from two biological replicate cultures for each condition.

#### *Exonuclease treatment*

The sample preparation for dRNAseq was accomplished essentially as described previously (Sharma et al., 2010). Briefly, total RNA extracted was divided in two equal portions, where one was incubated with Terminator TM 5'phosphate-dependent exonuclease (TEX) (Epicentre Illumina TER51020) to generate the primary transcript enriched library and the other left untreated. The exonuclease reaction was performed with 5 µg of total RNA sample, using 2 units of TEX for 1 h at 30°C. RiboLock RNase Inhibitor (Thermo Scientific E00381) was added to the reaction mixture (20 U) to preserve RNA integrity. The reaction was stopped with 1 µL of 100 mM EDTA. The RNA integrity and abundance of 16S and 23S rRNAs were assessed with a RNA 6000 Nano chip on the Agilent 2100 Bioanalyzer (Agilent Technologies) (Fig. S2).

TEX treatment was followed by phenol extraction and ethanol precipitation of mRNA, as described by the manufacture (Epicentre Illumina TER51020). RNase-free water was added to the reaction for a final volume of 200 µL. The extraction was performed once with an equal volume of buffer-saturated phenol, followed by vortexing and

centrifugation at (14500 x g, 2 min, 4°C). The aqueous phase was transferred to a new RNase-free tube, followed by addition of 1 mL precipitation mix (0.1 volume of 3 M sodium acetate pH 5.5, 2.5 volumes of cold ethanol 100%, 0.02 volume of glycogen 20 mg/μL (Thermo Scientific R0551)). After mixing thoroughly, the reaction was kept at -20°C for 30 min. The RNA was pelleted by centrifugation (14500 x g, 30 min, 4°C), and the supernatant discarded. The RNA pellet was washed with 500 μL of 70% ethanol and precipitated by centrifugation (14500 x g, 5 min, 4°C). The supernatant was discarded and the RNA pellet resuspended in 20 μL of RNase-free water. The final RNA samples were quantified using a NanoDrop 8000 (Thermo Scientific).

#### *Library preparation and RNA sequencing*

Sequencing libraries were constructed using the Illumina® TruSeq® Stranded mRNA Sample Preparation kit (Sultan et al., 2012). Each final library was validated with a DNA 1000 chip on the Agilent 2100 Bioanalyzer and concentration measured using a Qubit 2.0 Fluorometer (Invitrogen, Life Technologies). The concentration of each library was normalized to 10 nM using 10 mM Tris-Cl, pH 8.5, 0.1% Tween 20. Then, 10 μL of each normalized library were pooled together. The final pooled library sample was validated with the DNA High Sensitivity Assay on Agilent 2100 Bioanalyzer (Agilent Technologies) and the concentration confirmed on a Qubit 2.0 Fluorometer. The libraries were sequenced using the Illumina HiSeq2000 platform (Beckman Coulter Genomics).

### *Data analysis*

The sequencing reads were initially checked for quality by evaluation of average quality per reads Phred score and mapped onto the *P. putida* KT2440 genome (RefSeq Accession No. NC\_002947.3) with Bowtie2 (Langmead and Salzberg, 2012). Mapping output files were sorted and indexed with SAMtools (Li et al., 2009) and then converted to .wig files. The transcription start sites were identified by TSSpredator (Dugar et al., 2013) by processing the reads from biological replicate samples together and using the “more sensitivity” parameter settings, which determine TSS by *step height* and *processing site factor* values of 0.2 and 2, respectively. The assignment of primary and secondary TSSs was performed using the default value of a 300 nt maximal upstream distance from the start codon.

Statistical and data analysis were handled by R Bioconductor and Microsoft Excel. Promoter analysis were conducted by MEME Suite (Bailey et al., 2009) and *cis*-RNA secondary structures in 5'UTR regions were searched for homologies against the Rfam databases (Nawrocki et al., 2015). Operon and sRNA prediction were performed by Rockhopper (McClure et al., 2013; Tjaden, 2015). Visual inspection of identified putative sRNAs was done by the Integrative Genomics Viewer (IGV) (Robinson et al., 2011; Thorvaldsdottir et al., 2013). Novel ORFs were predicted by the *in silico* gene finders GLIMMER (Salzberg et al., 1998; Delcher et al., 2007) and GeneMark (Lukashin and Borodovsky, 1998; Besemer and Borodovsky, 2005) and the transcription confirmed by the RNA-seq data. ORFs functions were searched by sequence homology in protein Blast (Johnson et al., 2008).

### *5'RACE*

The 5' ends of mRNA transcripts were confirmed by RACE procedures published previously (Vogel and Wagner, 2005; Gomez-Lozano et al., 2012) with modifications. In our approach, the Tobacco Acid Pyrophosphatase treatment step of the DNase-treated total RNA was replaced by the TEX treatment (described above) followed by RNA 5'polyphosphatase (Epicentre Illumina RP8092H), which removes the  $\gamma$  and  $\beta$  phosphates from 5'triphosphorylated RNAs and has no activity on 5'monophosphorylated ends. Briefly, the untreated and treated TEX samples were incubated at 37°C with RNA 5'Polyphosphatase (20 Units) for 30 min. Following RNA purification, an RNA adapter was ligated to the 5'ends using T4 RNA ligase (Thermo Scientific EL0021). The adapter-RNA complex was reverse transcribed by Thermoscript RT-PCR System (Invitrogen 11146) using a gene specific primer (GSP1) for the mRNA. The resulting cDNA was amplified by PCR reaction with Phusion HotStart II High-Fidelity DNA polymerase (Thermo Scientific F-548S/L) using a second gene specific primer (GSP2) and an adapter-specific primer. The PCR products were checked on an agarose gel and sequenced with the PCR amplification primers at Eurofins Genomics (Denmark). The oligonucleotides used in this study are listed in Table S9.

### *Plasmid construction and uracil excision cloning*

PCR products were obtained using proof-reading PfuX7 polymerase (Norholm, 2010) in Phusion HF Buffer (Thermo Scientific). DNA fragments were amplified with 20 cycles in a 50  $\mu$ L reaction volume using a C1000 Touch<sup>TM</sup> Thermal Cyclers (BioRad). Amplicons were purified with PureLink<sup>TM</sup> Quick Gel Extraction and PCR Purification Combo Kit (Life Technologies) and quantified using NanoDrop 8000

(Thermo Scientific). Translational (pISO1) and transcriptional (pISO2) fusions of the TPP riboswitch and the reporter GFP in plasmid constructs were obtained via uracil excision cloning as described elsewhere (Cavaleiro et al., 2015a; Cavaleiro et al., 2015b). Oligonucleotides and plasmids used are listed in Table S9. Transformants were checked by colony PCR. Positive plasmids were isolated using the NucleoSpin R plasmid QuickPure Kit (Macherey-Nagel), sequenced at Eurofins Genomics (Denmark) and transformed into *P. putida* KT2440 (Martinez-Garcia and de Lorenzo, 2011).

#### *Riboswitch characterization*

Kinetic and manual assays for GFP fluorescence expression were performed using a SynergyMx 96-microtiter plate reader (BioTeck). Overnight cultures of bacteria grown in M9 medium with glucose (0.5% (w/v)) and gentamicin (10 µg/mL) were diluted to a starting OD<sub>600</sub> of 0.2 with fresh media and transferred to a 96 well microtiter plate. The plate was incubated at 30°C for two hours with shaking until an OD<sub>600</sub> of 0.2-0.3 was reached, followed by addition of the inducer 3-methylbenzoate at a final concentration of 0.5 mM. Kinetic assays were performed in the plate reader; where both cell density (OD<sub>600</sub>) and fluorescence (RFU (485, 528)) were measured every 10 min for *P. putida* KT2440 (pISO1) after induction, while manually measurement were performed every hour for *P. putida* KT2440 (pISO2). At 1.5 h after induction, ligands were added to the wells in different concentrations (10 nM, 100 nM, 1 µM, 5 µM, 10 µM, 100 µM, 0.5 mM, 1 mM, and 1.5 mM). Kinetic curves were monitored for 15 hours until the entry into stationary phase.



## **Acknowledgments**

The authors thank Ana Mafalda Cavaleiro for helpful discussions, Patricia Calero Valdayo for providing the plasmid pPCV31, and Holger Døssing for support and helpful comments on the manuscript. This work is supported by the Novo Nordisk Foundation Center for Biosustainability, and a PhD grant from the People Programme (Marie Curie Actions) of the European Union Seventh Framework Programme FP7-People-2012-ITN, under grant agreement No. 317058, “BACTORY”.

## **Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

## **References**

- Abril, M.A., Michan, C., Timmis, K.N., and Ramos, J.L. (1989) Regulator and enzyme specificities of the TOL plasmid-encoded upper pathway for degradation of aromatic hydrocarbons and expansion of the substrate range of the pathway. *J Bacteriol* **171**: 6782-6790.
- Araujo, P.R., Yoon, K., Ko, D., Smith, A.D., Qiao, M., Suresh, U. et al. (2012) Before It Gets Started: Regulating Translation at the 5' UTR. *Comp Funct Genomics* **2012**: 475731.
- Arnvig, K.B., and Young, D.B. (2009) Identification of small RNAs in *Mycobacterium tuberculosis*. *Mol Microbiol* **73**: 397-408.
- Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L. et al. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202-208.
- Begley, T.P., Downs, D.M., Ealick, S.E., McLafferty, F.W., Van Loon, A.P., Taylor, S. et al. (1999) Thiamin biosynthesis in prokaryotes. *Arch Microbiol* **171**: 293-300.

- Belda, E., van Heck, R.G., Lopez-Sanchez, M.J., Cruveiller, S., Barbe, V., Fraser, C. et al. (2016) The revisited genome of *Pseudomonas putida* KT2440 enlightens its value as a robust metabolic chassis. *Environ Microbiol.*
- Besemer, J., and Borodovsky, M. (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res* **33**: W451-454.
- Bischler, T., Tan, H.S., Nieselt, K., and Sharma, C.M. (2015) Differential RNA-seq (dRNA-seq) for annotation of transcriptional start sites and small RNAs in *Helicobacter pylori*. *Methods* **86**: 89-101.
- Blount, Z.D., Borland, C.Z., and Lenski, R.E. (2008) Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc Natl Acad Sci U S A* **105**: 7899-7906.
- Breaker, R.R. (2011) Prospects for riboswitch discovery and analysis. *Mol Cell* **43**: 867-879.
- Brock, J.E., Pourshahian, S., Giliberti, J., Limbach, P.A., and Janssen, G.R. (2008) Ribosomes bind leaderless mRNA in *Escherichia coli* through recognition of their 5'-terminal AUG. *RNA* **14**: 2159-2169.
- Caron, M.P., Bastet, L., Lussier, A., Simoneau-Roy, M., Masse, E., and Lafontaine, D.A. (2012) Dual-acting riboswitch control of translation initiation and mRNA decay. *Proc Natl Acad Sci U S A* **109**: E3444-3453.
- Cavaleiro, A.M., Kim, S.H., Seppala, S., Nielsen, M.T., and Norholm, M.H. (2015a) Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning. *ACS Synth Biol* **4**: 1042-1046.
- Cavaleiro, A.M., Nielsen, M.T., Kim, S.H., Seppälä, S., and Nørholm, M.H.H. (2015b) Uracil excision for assembly of complex pathways. In *Hydrocarbon and Lipid Microbiology Protocols*. McGenity, T.J., Timmis, K.N., and Nogales, B. (eds). Berlin Heidelberg: Springer-Verlag.
- Chavarria, M., Nikel, P.I., Perez-Pantoja, D., and de Lorenzo, V. (2013) The Entner-Doudoroff pathway empowers *Pseudomonas putida* KT2440 with a high tolerance to oxidative stress. *Environ Microbiol* **15**: 1772-1785.

Conway, T., Creecy, J.P., Maddox, S.M., Grissom, J.E., Conkle, T.L., Shadid, T.M. et al. (2014) Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. *MBio* **5**: e01442-01414.

Coppins, R.L., Hall, K.B., and Groisman, E.A. (2007) The intricate world of riboswitches. *Curr Opin Microbiol* **10**: 176-181.

Cortes, T., Schubert, O.T., Rose, G., Arnvig, K.B., Comas, I., Aebersold, R., and Young, D.B. (2013) Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*. *Cell Rep* **5**: 1121-1131.

Creecy, J.P., and Conway, T. (2015) Quantitative bacterial transcriptomics with RNA-seq. *Curr Opin Microbiol* **23**: 133-140.

Delcher, A.L., Bratke, K.A., Powers, E.C., and Salzberg, S.L. (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**: 673-679.

Dornenburg, J.E., Devita, A.M., Palumbo, M.J., and Wade, J.T. (2010) Widespread antisense transcription in *Escherichia coli*. *MBio* **1**.

Dugar, G., Herbig, A., Forstner, K.U., Heidrich, N., Reinhardt, R., Nieselt, K., and Sharma, C.M. (2013) High-resolution transcriptome maps reveal strain-specific regulatory features of multiple *Campylobacter jejuni* isolates. *PLoS Genet* **9**: e1003495.

Edwards, T.E., and Ferre-D'Amare, A.R. (2006) Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure* **14**: 1459-1468.

Federal Register (1982) Appendix E, Certified host-vector systems. **47**: 17197.

Filiatrault, M.J., Stodghill, P.V., Myers, C.R., Bronstein, P.A., Butcher, B.G., Lam, H. et al. (2011) Genome-wide identification of transcriptional start sites in the plant pathogen *Pseudomonas syringae* pv. tomato str. DC3000. *PLoS One* **6**: e29335.

Filiatrault, M.J., Stodghill, P.V., Bronstein, P.A., Moll, S., Lindeberg, M., Grills, G. et al. (2010) Transcriptome analysis of *Pseudomonas syringae*

identifies new genes, noncoding RNAs, and antisense activity. *J Bacteriol* **192**: 2359-2372.

Frank, S., Klockgether, J., Hagendorf, P., Geffers, R., Schock, U., Pohl, T. et al. (2011) *Pseudomonas putida* KT2440 genome update by cDNA sequencing and microarray transcriptomics. *Environ Microbiol* **13**: 1309-1326.

Georg, J., and Hess, W.R. (2011) *cis*-antisense RNA, another level of gene regulation in bacteria. *Microbiol Mol Biol Rev* **75**: 286-300.

Georg, J., Voss, B., Scholz, I., Mitschke, J., Wilde, A., and Hess, W.R. (2009) Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. *Mol Syst Biol* **5**: 305.

Gomez-Lozano, M., Marvig, R.L., Molin, S., and Long, K.S. (2012) Genome-wide identification of novel small RNAs in *Pseudomonas aeruginosa*. *Environ Microbiol* **14**: 2006-2016.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R., and Hofacker, I.L. (2008) The Vienna RNA websuite. *Nucleic Acids Res* **36**: W70-74.

Guell, M., van Noort, V., Yus, E., Chen, W.H., Leigh-Bell, J., Michalodimitrakis, K. et al. (2009) Transcriptome complexity in a genome-reduced bacterium. *Science* **326**: 1268-1271.

Haller, A., Altman, R.B., Souliere, M.F., Blanchard, S.C., and Micura, R. (2013) Folding and ligand recognition of the TPP riboswitch aptamer at single-molecule resolution. *Proc Natl Acad Sci U S A* **110**: 4188-4193.

Inoue, H., Nojima, H., and Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**: 23-28.

Irnov, I., Sharma, C.M., Vogel, J., and Winkler, W.C. (2010) Identification of regulatory RNAs in *Bacillus subtilis*. *Nucleic Acids Res* **38**: 6637-6651.

Jeanguenin, L., Lara-Nunez, A., Rodionov, D.A., Osterman, A.L., Komarova, N.Y., Rentsch, D. et al. (2012) Comparative genomics and functional analysis of the NiaP family uncover nicotinate transporters from bacteria, plants, and mammals. *Funct Integr Genomics* **12**: 25-34.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res* **36**: W5-9.

Jurgenson, C.T., Begley, T.P., and Ealick, S.E. (2009) The structural and biochemical foundations of thiamin biosynthesis. *Annu Rev Biochem* **78**: 569-603.

Kim, J., Oliveros, J.C., Nickel, P.I., de Lorenzo, V., and Silva-Rocha, R. (2013) Transcriptomic fingerprinting of *Pseudomonas putida* under alternative physiological regimes. *Environ Microbiol Rep* **5**: 883-891.

Kopf, M., and Hess, W.R. (2015) Regulatory RNAs in photosynthetic cyanobacteria. *FEMS Microbiol Rev* **39**: 301-315.

Kopf, M., Klahn, S., Scholz, I., Hess, W.R., and Voss, B. (2015) Variations in the non-coding transcriptome as a driver of inter-strain divergence and physiological adaptation in bacteria. *Sci Rep* **5**: 9560.

Kopf, M., Klahn, S., Pade, N., Weingartner, C., Hagemann, M., Voss, B., and Hess, W.R. (2014) Comparative genome analysis of the closely related *Synechocystis* strains PCC 6714 and PCC 6803. *DNA Res* **21**: 255-266.

Kroger, C., Dillon, S.C., Cameron, A.D., Papenfort, K., Sivasankaran, S.K., Hokamp, K. et al. (2012) The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc Natl Acad Sci U S A* **109**: E1277-1286.

La Rosa, R., Nogales, J., and Rojo, F. (2015a) The Crc/CrcZ-CrcY global regulatory system helps the integration of gluconeogenic and glycolytic metabolism in *Pseudomonas putida*. *Environ Microbiol* **17**: 3362-3378.

La Rosa, R., Behrends, V., Williams, H.D., Bundy, J.G., and Rojo, F. (2015b) Influence of the Crc regulator on the hierarchical use of carbon sources from a complete medium in *Pseudomonas*. *Environ Microbiol*.

Lang, K., Rieder, R., and Micura, R. (2007) Ligand-induced folding of the *thiM* TPP riboswitch investigated by a structure-based fluorescence spectroscopic approach. *Nucleic Acids Res* **35**: 5370-5378.

Langmead, B., and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.

Laursen, B.S., Sorensen, H.P., Mortensen, K.K., and Sperling-Petersen, H.U. (2005) Initiation of protein synthesis in bacteria. *Microbiol Mol Biol Rev* **69**: 101-123.

Lawrence, J.G. (2002) Shared strategies in gene organization among prokaryotes and eukaryotes. *Cell* **110**: 407-413.

Lehmann, K., and Schmidt, U. (2003) Group II introns: structure and catalytic versatility of large natural ribozymes. *Crit Rev Biochem Mol Biol* **38**: 249-303.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Liu, J.M., Livny, J., Lawrence, M.S., Kimball, M.D., Waldor, M.K., and Camilli, A. (2009) Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res* **37**: e46.

Loeschcke, A., and Thies, S. (2015) *Pseudomonas putida*-a versatile host for the production of natural products. *Appl Microbiol Biotechnol* **99**: 6197-6214.

Lugtenberg, B.J., Dekkers, L., and Bloemberg, G.V. (2001) Molecular determinants of rhizosphere colonization by *Pseudomonas*. *Annu Rev Phytopathol* **39**: 461-490.

Lukashin, A.V., and Borodovsky, M. (1998) GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res* **26**: 1107-1115.

Lunse, C.E., Schuller, A., and Mayer, G. (2014) The promise of riboswitches as potential antibacterial drug targets. *Int J Med Microbiol* **304**: 79-92.

Mandin, P., Repoila, F., Vergassola, M., Geissmann, T., and Cossart, P. (2007) Identification of new noncoding RNAs in *Listeria monocytogenes* and prediction of mRNA targets. *Nucleic Acids Res* **35**: 962-974.

Marques, S., Gallegos, M.T., Manzanera, M., Holtel, A., Timmis, K.N., and Ramos, J.L. (1998) Activation and repression of transcription at the double tandem divergent promoters for the *xylR* and *xylS* genes of the TOL plasmid of *Pseudomonas putida*. *J Bacteriol* **180**: 2889-2894.

Martinez-Garcia, E., and de Lorenzo, V. (2011) Engineering multiple genomic deletions in Gram-negative bacteria: analysis of the multi-resistant antibiotic profile of *Pseudomonas putida* KT2440. *Environ Microbiol* **13**: 2702-2716.

McClure, R., Balasubramanian, D., Sun, Y., Bobrovskyy, M., Sumby, P., Genco, C.A. et al. (2013) Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res* **41**: e140.

Miranda-Rios, J. (2007) The THI-box riboswitch, or how RNA binds thiamin pyrophosphate. *Structure* **15**: 259-265.

Mironov, A.S., Gusarov, I., Rafikov, R., Lopez, L.E., Shatalin, K., Kreneva, R.A. et al. (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* **111**: 747-756.

Mitschke, J., Vioque, A., Haas, F., Hess, W.R., and Muro-Pastor, A.M. (2011) Dynamics of transcriptional start site selection during nitrogen stress-induced cell differentiation in *Anabaena* sp. PCC7120. *Proc Natl Acad Sci U S A* **108**: 20130-20135.

Miura, K., Inouye, S., and Nakazawa, A. (1998) The *rpoS* gene regulates OP2, an operon for the lower pathway of xylene catabolism on the TOL plasmid, and the stress response in *Pseudomonas putida* mt-2. *Mol Gen Genet* **259**: 72-78.

Moll, I., Grill, S., Gualerzi, C.O., and Blasi, U. (2002) Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. *Mol Microbiol* **43**: 239-246.

Munoz-Elias, E.J., and McKinney, J.D. (2006) Carbon metabolism of intracellular bacteria. *Cell Microbiol* **8**: 10-22.

Nakagawa, S., Niimura, Y., Miura, K., and Gojobori, T. (2010) Dynamic evolution of translation initiation mechanisms in prokaryotes. *Proc Natl Acad Sci U S A* **107**: 6382-6387.

Nakazawa, T. (2002) Travels of a *Pseudomonas*, from Japan around the world. *Environ Microbiol* **4**: 782-786.

Narlikar, L. (2014) Multiple novel promoter-architectures revealed by decoding the hidden heterogeneity within the genome. *Nucleic Acids Res* **42**: 12388-12403.

Naville, M., and Gautheret, D. (2010) Premature terminator analysis sheds light on a hidden world of bacterial transcriptional attenuation. *Genome Biol* **11**: R97.

Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., Eberhardt, R.Y., Eddy, S.R. et al. (2015) Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res* **43**: D130-137.

Nelson, K.E., Weinel, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Martins dos Santos, V.A. et al. (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol* **4**: 799-808.

Nikel, P.I., Martinez-Garcia, E., and de Lorenzo, V. (2014) Biotechnological domestication of pseudomonads using synthetic biology. *Nat Rev Microbiol* **12**: 368-379.

Nogales, J., Palsson, B.O., and Thiele, I. (2008) A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iJN746 as a cell factory. *BMC Syst Biol* **2**: 79.

Norholm, M.H. (2010) A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. *BMC Biotechnol* **10**: 21.

Nudler, E., and Mironov, A.S. (2004) The riboswitch control of bacterial metabolism. *Trends Biochem Sci* **29**: 11-17.

Nuss, A.M., Heroven, A.K., Waldmann, B., Reinkensmeier, J., Jarek, M., Beckstette, M., and Dersch, P. (2015) Transcriptomic profiling of *Yersinia pseudotuberculosis* reveals reprogramming of the Crp regulon by temperature and uncovers Crp as a master regulator of small RNAs. *PLoS Genet* **11**: e1005087.

Oberhardt, M.A., Puchalka, J., Martins dos Santos, V.A., and Papin, J.A. (2011) Reconciliation of genome-scale metabolic reconstructions for comparative systems analysis. *PLoS Comput Biol* **7**: e1001116.

Ontiveros-Palacios, N., Smith, A.M., Grundy, F.J., Soberon, M., Henkin, T.M., and Miranda-Rios, J. (2008) Molecular basis of gene regulation by the THI-box riboswitch. *Mol Microbiol* **67**: 793-803.

Opdyke, J.A., Kang, J.G., and Storz, G. (2004) GadY, a small-RNA regulator of acid response genes in *Escherichia coli*. *J Bacteriol* **186**: 6698-6705.

Papenfort, K., Forstner, K.U., Cong, J.P., Sharma, C.M., and Bassler, B.L. (2015) Differential RNA-seq of *Vibrio cholerae* identifies the VqmR small



RNA as a regulator of biofilm formation. *Proc Natl Acad Sci U S A* **112**: E766-775.

Pfreundt, U., Kopf, M., Belkin, N., Berman-Frank, I., and Hess, W.R. (2014) The primary transcriptome of the marine diazotroph *Trichodesmium erythraeum* IMS101. *Sci Rep* **4**: 6187.

Poblete-Castro, I., Becker, J., Dohnt, K., dos Santos, V.M., and Wittmann, C. (2012) Industrial biotechnology of *Pseudomonas putida* and related species. *Appl Microbiol Biotechnol* **93**: 2279-2290.

Puchalka, J., Oberhardt, M.A., Godinho, M., Bielecka, A., Regenhardt, D., Timmis, K.N. et al. (2008) Genome-scale reconstruction and analysis of the *Pseudomonas putida* KT2440 metabolic network facilitates applications in biotechnology. *PLoS Comput Biol* **4**: e1000210.

Ramesh, A. (2015) Second messenger - Sensing riboswitches in bacteria. *Semin Cell Dev Biol* **47-48**: 3-8.

Regenhardt, D., Heuer, H., Heim, S., Fernandez, D.U., Strompl, C., Moore, E.R., and Timmis, K.N. (2002) Pedigree and taxonomic credentials of *Pseudomonas putida* strain KT2440. *Environ Microbiol* **4**: 912-915.

Revelles, O., Espinosa-Urgel, M., Fuhrer, T., Sauer, U., and Ramos, J.L. (2005) Multiple and interconnected pathways for L-lysine catabolism in *Pseudomonas putida* KT2440. *J Bacteriol* **187**: 7500-7510.

Righetti, F., and Narberhaus, F. (2014) How to find RNA thermometers. *Front Cell Infect Microbiol* **4**: 132.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011) Integrative genomics viewer. *Nat Biotechnol* **29**: 24-26.

Rodionov, D.A., Hebbeln, P., Eudes, A., ter Beek, J., Rodionova, I.A., Erkens, G.B. et al. (2009) A novel class of modular transporters for vitamins in prokaryotes. *J Bacteriol* **191**: 42-51.

Rodionova, I.A., Li, X., Plymale, A.E., Motamedchaboki, K., Konopka, A.E., Romine, M.F. et al. (2015) Genomic distribution of B-vitamin auxotrophy and uptake transporters in environmental bacteria from the *Chloroflexi* phylum. *Environ Microbiol Rep* **7**: 204-210.

Rojo, F. (2010) Carbon catabolite repression in *Pseudomonas* : optimizing metabolic versatility and interactions with the environment. *FEMS Microbiol Rev* **34**: 658-684.

Sahr, T., Rusniok, C., Dervins-Ravault, D., Sismeiro, O., Coppee, J.Y., and Buchrieser, C. (2012) Deep sequencing defines the transcriptional map of *L. pneumophila* and identifies growth phase-dependent regulated ncRNAs implicated in virulence. *RNA Biol* **9**: 503-519.

Salzberg, S.L., Delcher, A.L., Kasif, S., and White, O. (1998) Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* **26**: 544-548.

Schlax, P.J., Xavier, K.A., Gluick, T.C., and Draper, D.E. (2001) Translational repression of the *Escherichia coli* alpha operon mRNA: importance of an mRNA conformational switch and a ternary entrapment complex. *J Biol Chem* **276**: 38494-38501.

Schluter, J.P., Reinkensmeier, J., Barnett, M.J., Lang, C., Krol, E., Giegerich, R. et al. (2013) Global mapping of transcription start sites and promoter motifs in the symbiotic alpha-proteobacterium *Sinorhizobium meliloti* 1021. *BMC Genomics* **14**: 156.

Schluter, J.P., Reinkensmeier, J., Daschkey, S., Evguenieva-Hackenberg, E., Janssen, S., Janicke, S. et al. (2010) A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium *Sinorhizobium meliloti*. *BMC Genomics* **11**: 245.

Schmidtke, C., Findeiss, S., Sharma, C.M., Kuhfuss, J., Hoffmann, S., Vogel, J. et al. (2012) Genome-wide transcriptome analysis of the plant pathogen *Xanthomonas* identifies sRNAs with putative virulence functions. *Nucleic Acids Res* **40**: 2020-2031.

Sharma, C.M., and Vogel, J. (2014) Differential RNA-seq: the approach behind and the biological insight gained. *Curr Opin Microbiol* **19**: 97-105.

Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A. et al. (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464**: 250-255.

Shell, S.S., Wang, J., Lapierre, P., Mir, M., Chase, M.R., Pyle, M.M. et al. (2015) Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape. *PLoS Genet* **11**: e1005641.

Silby, M.W., and Levy, S.B. (2008) Overlapping protein-encoding genes in *Pseudomonas fluorescens* Pf0-1. *PLoS Genet* **4**: e1000094.

Sohn, S.B., Kim, T.Y., Park, J.M., and Lee, S.Y. (2010) In silico genome-scale metabolic analysis of *Pseudomonas putida* KT2440 for polyhydroxyalkanoate synthesis, degradation of aromatics and anaerobic survival. *Biotechnol J* **5**: 739-750.

Sultan, M., Dokel, S., Amstislavskiy, V., Wuttig, D., Sultmann, H., Lehrach, H., and Yaspo, M.L. (2012) A simple strand-specific RNA-Seq library preparation protocol combining the Illumina TruSeq RNA and the dUTP methods. *Biochem Biophys Res Commun* **422**: 643-646.

Sun, E.I., Leyn, S.A., Kazanov, M.D., Saier, M.H., Jr., Novichkov, P.S., and Rodionov, D.A. (2013) Comparative genomics of metabolic capacities of regulons controlled by *cis*-regulatory RNA motifs in bacteria. *BMC Genomics* **14**: 597.

Thomason, M.K., Bischler, T., Eisenbart, S.K., Forstner, K.U., Zhang, A., Herbig, A. et al. (2015) Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. *J Bacteriol* **197**: 18-28.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178-192.

Tjaden, B. (2015) *De novo* assembly of bacterial transcriptomes from RNA-seq data. *Genome Biol* **16**: 1.

Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D. et al. (2009) The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature* **459**: 950-956.

Vilchez, S., Molina, L., Ramos, C., and Ramos, J.L. (2000) Proline catabolism by *Pseudomonas putida*: cloning, characterization, and expression of the *put* genes in the presence of root exudates. *J Bacteriol* **182**: 91-99.

Vogel, J., and Wagner, E.G.H. (2005) Approaches to identify novel non-messenger RNAs in bacteria and to investigate their biological functions: RNA mining. In *Handbook of RNA Biochemistry*. Hartmann, R.K., Bindereif, A., Schön, A., and Westhof, E. (eds). Weinham, Germany: Wiley-VCH Verlag GmbH.

Voigt, K., Sharma, C.M., Mitschke, J., Lambrecht, S.J., Voss, B., Hess, W.R., and Steglich, C. (2014) Comparative transcriptomics of two environmentally relevant cyanobacteria reveals unexpected transcriptome diversity. *ISME J* **8**: 2056-2068.

Voss, B., Bolhuis, H., Fewer, D.P., Kopf, M., Moke, F., Haas, F. et al. (2013) Insights into the physiology and ecology of the brackish-water-adapted Cyanobacterium *Nodularia spumigena* CCY9414 based on a genome-transcriptome analysis. *PLoS One* **8**: e60224.

Waters, L.S., and Storz, G. (2009) Regulatory RNAs in bacteria. *Cell* **136**: 615-628.

Webb, E., Claas, K., and Downs, D. (1998) *thiBPQ* encodes an ABC transporter required for transport of thiamine and thiamine pyrophosphate in *Salmonella typhimurium*. *J Biol Chem* **273**: 8946-8950.

Weinberg, Z., Wang, J.X., Bogue, J., Yang, J., Corbino, K., Moy, R.H., and Breaker, R.R. (2010) Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol* **11**: R31.

Weinberg, Z., Barrick, J.E., Yao, Z., Roth, A., Kim, J.N., Gore, J. et al. (2007) Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res* **35**: 4809-4819.

Wiegand, S., Dietrich, S., Hertel, R., Bongaerts, J., Evers, S., Volland, S. et al. (2013) RNA-Seq of *Bacillus licheniformis*: active regulatory RNA features expressed within a productive fermentation. *BMC Genomics* **14**: 667.

Winkler, W., Nahvi, A., and Breaker, R.R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**: 952-956.

Winkler, W.C., and Breaker, R.R. (2003) Genetic control by metabolite-binding riboswitches. *ChemBiochem* **4**: 1024-1032.

Winther-Larsen, H.C., Josefsen, K.D., Brautaset, T., and Valla, S. (2000) Parameters affecting gene expression from the Pm promoter in gram-negative bacteria. *Metab Eng* **2**: 79-91.

Wurtzel, O., Sesto, N., Mellin, J.R., Karunker, I., Edelheit, S., Becavin, C. et al. (2012) Comparative transcriptomics of pathogenic and non-pathogenic *Listeria* species. *Mol Syst Biol* **8**: 583.

Yamauchi, T., Miyoshi, D., Kubodera, T., Nishimura, A., Nakai, S., and Sugimoto, N. (2005) Roles of Mg<sup>2+</sup> in TPP-dependent riboswitch. *FEBS Lett* **579**: 2583-2588.

Zheng, X., Hu, G.Q., She, Z.S., and Zhu, H. (2011) Leaderless genes in bacteria: clue to the evolution of translation initiation mechanisms in prokaryotes. *BMC Genomics* **12**: 361.

## **Supplementary Information**

**Figure S1.** Growth curves of *P. putida* KT2440.

**Figure S2.** Size profiles of total and exonuclease-treated RNA samples.

**Figure S3.** Comparison of TSS classification between differential RNA-seq studies in other bacteria.

**Figure S4.** 5' RACE experiment to confirm predicted TSSs.

**Figure S5.** Overview of multi-gene operons identified in *P. putida* KT2440.

**Figure S6.** Profiles of transcript categories.

**Table S1.** Mapping statistics.

**Table S2.** Transcription start sites predicted by TSSpredator. (Not reported in this thesis as too long).

**Table S3.** Summary of 5' RACE results.

**Table S4.** Multi-gene operons. (Not reported in this thesis as too long).

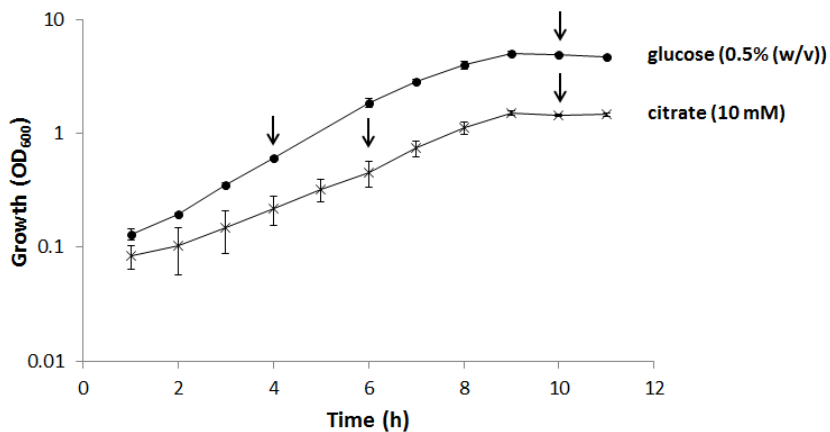
**Table S5.** Leaderless mRNAs.

**Table S6.** Intergenic small RNA transcripts.

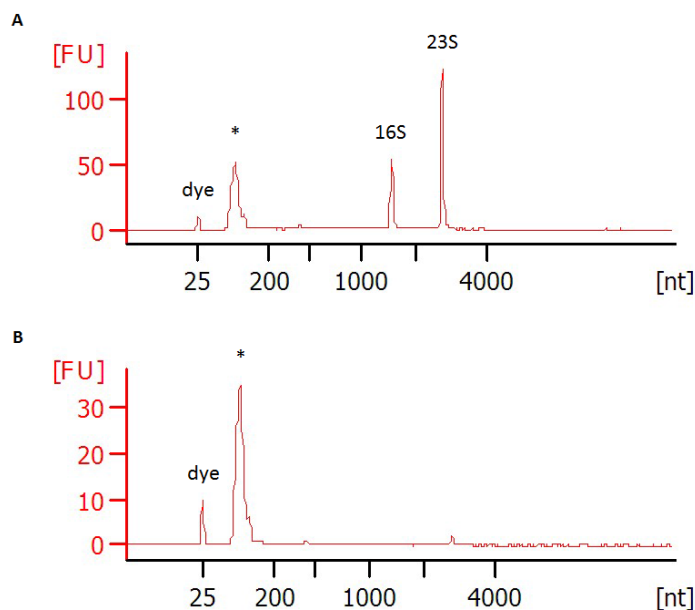
**Table S7.** Putative actuatons.

**Table S8.** Putative ORFs.

**Table S9.** Strains, plasmids and oligonucleotides used in this work.

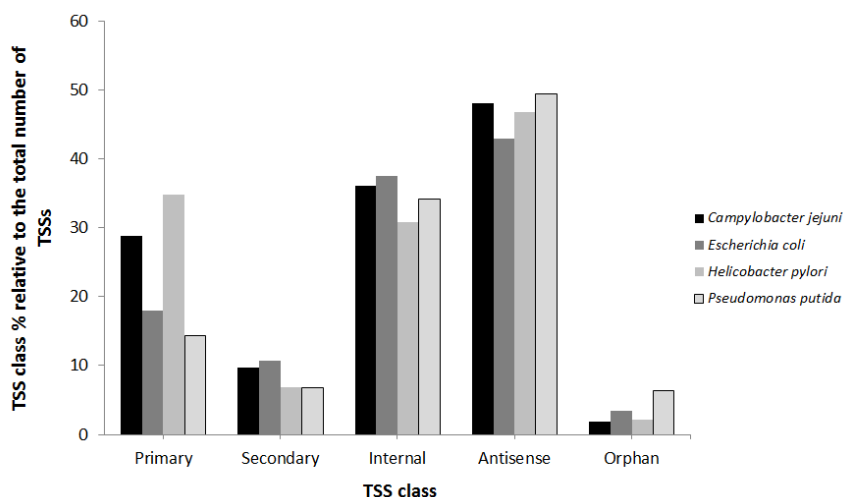


**Figure S1.** Growth curves of *P. putida* KT2440. Cells were grown in M9 minimal medium in the presence of glucose (0.5% (w/v)) or citrate (10 mM) as sole carbon sources. Cell growth was carried out at 30° C with shaking at 250 rpm, and monitored by measuring optical density at 600 nm. Growth rate in glucose  $\mu_{\text{MAX}}$  ( $\text{h}^{-1}$ )  $0.55 \pm 0.01$ , growth rate in citrate  $\mu_{\text{MAX}}$  ( $\text{h}^{-1}$ )  $0.40 \pm 0.05$ . Arrows indicate the cell harvest points. Error bars represent the standard deviations of three biological replicates.

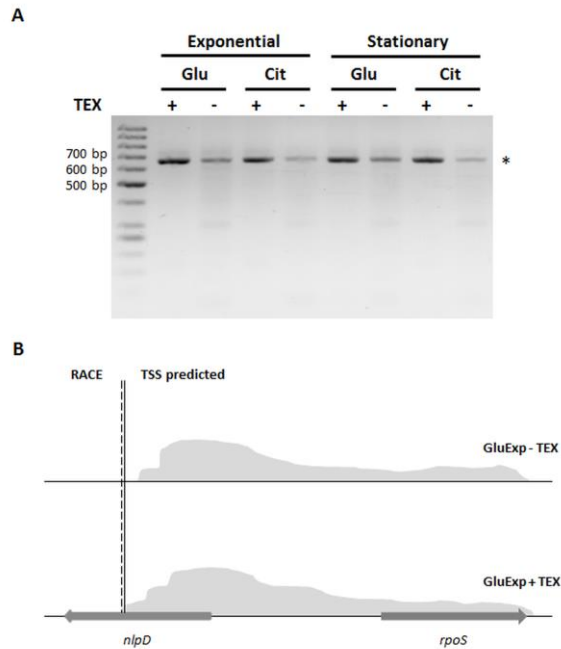


**Figure S2.** Size profiles of total and exonuclease-treated RNA samples. RNA extracted from *P. putida* KT2440 cells was analyzed with the RNA 6000 Nano chip on the Bioanalyzer (Agilent). Example profiles of total RNA (A) and exonuclease-treated RNA (B) samples. The peaks corresponding to 16S and 23S rRNAs are labeled (16S, 23S). The peaks marked with asterisks (\*) correspond to 5S rRNA, tRNAs and RNA transcripts shorter than 120 nt. The absence of 16S and 23S rRNA peaks in (B) shows the effect of the exonuclease enzyme on processed RNAs.



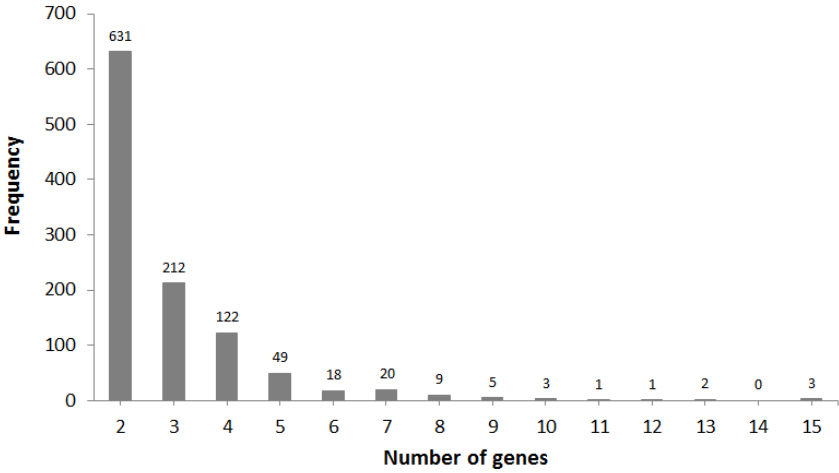


**Figure S3.** Comparison of TSS classification between differential RNA-seq studies in other bacteria. The *P. putida* KT2440 TSS classification in this study is compared with that of previous studies in *Campylobacter jejuni* (Dugar et al., 2013), *Escherichia coli* (Thomason et al., 2015) and *Helicobacter pylori* (Bischler et al., 2015). All studies used exonuclease treatment, the software TSSpredator for TSS identification and the same TSS class definitions.

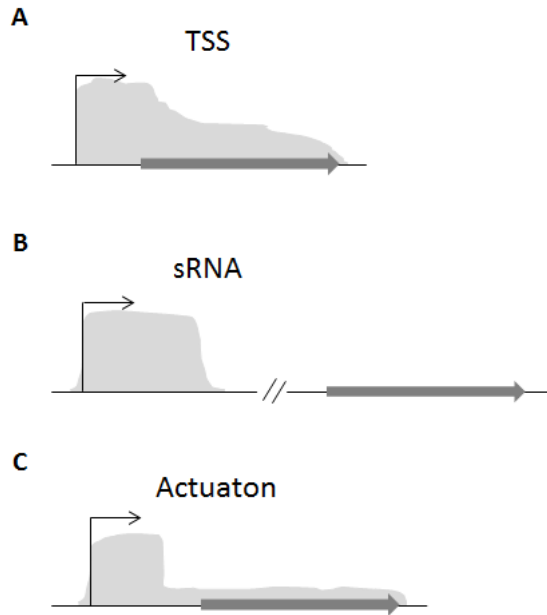


**Figure S4.** 5'RACE experiment to confirm predicted TSSs. A 5'RACE experiment was performed on the *rpoS* transcript (PP1623) in all growth conditions (GluExp, CitExp, GluSta, CitSta). Total RNA samples were divided into two equal parts where one was treated with exonuclease enzyme (+) and the other left untreated (-). Both samples were treated with RNA polyphosphatase, which removes the two phosphate groups at the 5'ends of primary RNA transcripts. Then the RNA adapter was ligated to the 5'ends, followed by reverse transcription performed with the gene specific primer that is specific for the *rpoS* transcripts. The cDNA was amplified by PCR with a second specific primer. A. The 2% agarose gel shows the primary bands (\*) after RACE that were sequenced. An enriched band corresponding to the primary transcript is observed for the '+' samples, whereas the '-' samples contain both primary and shorter processed transcripts. B. Schematic representation of the reads profile surrounding the *rpoS* gene in the glucose exponential growth condition. The

picture shows the mapped reads to the *rpoS* and the upstream *nlpD* genes (PP1622) in both untreated (top) and treated (bottom) samples. The treated sample shows a higher number of reads corresponding to the primary transcript comparing with the untreated sample. The TSSpredator identified TSS coordinates (solid line) internal to *nlpD* gene. The 5'RACE results (dashed line) show the point where the TSS has been determined by the 5'RACE experiment. There is a difference of only 4 nt between the predicted and RACE determined TSS positions.



**Figure S5.** Overview of multi-gene operons identified in *P. putida* KT2440. A total of 1076 multi-gene operons were found in *P. putida* KT2440. The graph summarizes the operons based on the number of genes they contain.



**Figure S6.** Profiles of transcript categories. The three different profiles associated with particular transcript categories detected in this study are illustrated, where arrows denote TSSs. A. Read profile of the TSS associated with an expressed mRNA, where there is a high intensity of reads mapping to the 5'UTR and expression that continues through the gene. B. Typical sRNA profile with high expression in an intergenic region. C. Actuation profile with a high number of reads just downstream of the TSS in the 5'UTR, followed by low or no expression extending through the downstream gene.

**Table S1. Mapping statistics.**

The total number of reads per sample obtained by dRNA-seq and their mapping on the *Pseudomonas putida* KT2440 genome. The number of reads mapping to rRNA and unique locations are also reported.

Sample	CitExp1	ECitExp1	CitExp2	CitSta1	ECitSta1	CitSta2	ECitSta2
Total reads	14606687	7319772	2467876	10773486	13792053	2529327	4425463
Mapped reads (%)	13990359 (96%)	7216017 (99%)	2239434 (91%)	10493349 (97%)	13557678 (98%)	2371249 (94%)	4346004 (98%)
rRNA reads (%)	13570648 (97%)	5772814 (80%)	2172251 (97%)	10283482 (98%)	11930757 (88%)	2323824 (98%)	3433343 (79%)
Unique reads (%)	419711 (3%)	1371043 (19%)	89577 (4%)	314800 (3%)	1626921 (12%)	47425 (2%)	912661 (21%)

Sample	GluExp1	EGluExp1	GluExp2	EGluExp2	GluSta1	EGluSta1	GluSta2	EGluSta2
Total reads	4979158	8498168	7357738	6214340	5662585	9756498	10872646	7362417
Mapped reads (%)	4618726 (93%)	8273117 (97%)	7112066 (97%)	6087137 (98%)	5466349 (97%)	9466666 (97%)	10620757 (98%)	7199819 (98%)
rRNA reads (%)	4341602 (94%)	6783956 (82%)	6756463 (95%)	4747967 (78%)	5302359 (97%)	8803999 (93%)	10408342 (98%)	6047848 (84%)
Unique reads (%)	277124 (6%)	1489161 (18%)	355603 (5%)	1217427 (20%)	163990 (3%)	662667 (7%)	212415 (2%)	1079973 (15%)

\* The % of mapped reads relative to the total number of reads.

\*\* The % of reads relative to the total number of mapped reads.

Abbreviations: Cit, citrate; Glu, glucose; Exp, exponential phase; Sta, stationary phase; E, exonuclease-treated sample.

**Table S3. Summary of 5'RACE results.**

RACE was performed on three genes in both treated and untreated conditions.

The PP designations of the genes, strand location, tested conditions, as well as the TSS coordinates predicted by TSS predator, determined by 5'RACE, and the difference between them are given.

Sample number	Gene number	Gene designation	Strand	Condition tested	Predicted TSS	5' RACE TSS	Difference (nt)
1	PP0147	citrate transporter	-	CitExp	157929	157923	6
2	PP4010	<i>cspD</i>	+	CitSta	4520037	4520028	9
3	PP4010	<i>cspD</i>	+	GluSta	4520037	4520028	9
4	PP1623	<i>rpoS</i>	+	CitExp	1818387	1818383	4
5	PP1623	<i>rpoS</i>	+	GluExp	1818387	1818383	4
6	PP1623	<i>rpoS</i>	+	CitSta	1818387	1818383	4
7	PP1623	<i>rpoS</i>	+	GluSta	1818387	1818383	4

Abbreviations: Cit, citrate; Glu, glucose, Exp, exponential phase; Sta, stationary phase.

**Table S5. Leaderless mRNAs.**

A total of 51 leaderless mRNAs have been identified.

For each leaderless mRNA, the genomic coordinate of start position, strand, locus information (locus designation, gene product, gene length), and UTR length are reported.

Position	Strand	Locus tag	Product	Gene length	UTR length
72853	-	PP0061	glycyl-tRNA synthetase subunit alpha	948	5
475702	+	PP0390	DNA-binding/iron metalloprotein/AP endonuclease	1026	6
605646	+	PP0520	phosphatidylglycerophosphatase A	504	5
740838	+	PP0634	fimbrial protein pilin	411	3
874415	+	PP0759	hypothetical protein	840	5
1242472	-	PP1082	bacterioferritin	474	8
1242889	-	PP1083	BFD (2Fe-2S)-binding domain-containing protein	219	7
1257452	+	PP1100	deoxycytidine triphosphate deaminase	567	4
1311534	-	PP1144	diguanylate cyclase	2265	4
1390189	+	PP1213	aspartyl-tRNA synthetase	1776	3
1417530	+	PP1240	phosphoribosylaminoimidazolesuccinocarboxamide synthase	711	10
1733866	-	PP1526	beta-(1-3)-glucosyl transferase	2592	1
1778150	-	PP1586	killer protein	279	10
1799198	+	PP1605	ribonuclease HII	624	5
1848550	+	PP1654	cysteine synthase B	900	6
2080503	+	PP1858	elongation factor P	570	7
2227402	-	PP1964	deoxynucleotide monophosphate kinase	804	4
2380327	+	PP2089	OmpF family protein	1035	7
2478293	+	PP2172	hypothetical protein	444	8
2624401	+	PP2296	hypothetical protein	357	5
2882011	-	PP2536	glutathione S-transferase	624	3
3282918	-	PP2884	XRE family transcriptional regulator	546	8

3328446	+	PP2927	hypothetical protein	1047	1
3727917	-	PP3293	hypothetical protein	426	3
4127622	+	PP3631	hypothetical protein	651	6
4148759	-	PP3652	amino acid transporter LysE	633	7
4257617	-	PP3730	transcriptional regulator	717	1
4266462	-	PP3738	GntR family transcriptional regulator	714	7
4361547	-	PP3836	hypothetical protein	336	8
4460531	+	PP3954	hypothetical protein	951	3
4519523	-	PP4008	ATP-dependent Clp protease ATP-binding subunit ClpA	2271	5
4547348	-	PP4035	NCS1 nucleoside transporter	1491	6
4676114	+	PP4139	hypothetical protein	219	5
4873618	-	PP4282	aquaporin Z	693	0
4936996	+	PP4345	GntR family transcriptional regulator	693	10
5081537	-	PP4473	aspartate kinase	1236	1
5120350	+	PP4507	TrkH family potassium uptake protein	1455	6
5143783	-	PP4527	hypothetical protein	927	2
5210142	-	PP4591	ribonuclease D	1134	10
5270336	-	PP4644	DNA repair protein RadA	1371	0
5431242	-	PP4770	periplasmic ligand-binding sensor protein	759	7
5451465	+	PP4790	apolipoprotein N-acyltransferase	1518	3
5741325	-	PP5038	hypothetical protein	255	5
5745862	-	PP5041	glycogen/starch/alpha-glucan phosphorylase	2451	7
5904741	-	PP5177	ornithine carbamoyltransferase	894	6
5936466	+	PP5206	secretion protein HlyD family protein	960	9
6033659	+	PP5285	bifunctional phosphopantothencysteine decarboxylase/phosphopantothenate synthase	1212	10
6104119	+	PP5354	hypothetical protein	408	7
6108146	-	PP5357	pyridoxamine kinase	873	4
6147816	-	PP5393	heavy metal transport/detoxification protein	201	10
6175174	-	PP5412	ATP synthase F0F1 subunit epsilon	420	8



**Table S6. Intergenic small RNA transcripts.**

A total of 80 small RNAs candidates have been identified and are listed according to their genomic coordinates.

For each transcript the name, start and stop coordinates, length, strand, 5' and 3' flanking genes, and orientation relative to the flanking genes are indicated.

Number	Name	Start	Stop	Length	Strand	5' Flanking gene	3' Flanking gene	Orientation
1	Pit003	16329	16281	49	-	PP0013	PP0014	><>
2	Pit006	58407	58555	149	+	PP0049	PP0050	<><
3	Spot42-like/spf/ErsA*	130367	130539	173	+	PP0123	PP0124	>><
4	C4 AS RNA 1*	335696	335870	175	+	PP0277	PP0278	<>>
5	RsmY*	450782	450934	153	+	PP0370	PP0371	>><
6	Pit020	450917	450814	104	-	PP0370	PP0371	><<
7	P27*	536446	536303	144	-	PP0444	PP0445	><>
8	P26*	537405	537502	98	+	PP0446	PP0447	>>>
9	Pit024	611076	610907	170	-	PP0525	PP0526	<<>
10	Pit025	624099	624004	96	-	PP0536	PP0537	<<<
11	Pit031	751928	752255	328	+	PP0640	PP0641	<><
12	C4 AS RNA 2*	759558	759654	97	-	PP0651	PP0652	>><
13	Pit048	1296778	1296615	164	-	PP1132	PP1133	<<>
14	Pit049	1298345	1298507	163	+	PP1132	PP1133	>><
15	PhrS*	1316300	1316400	101	+	PP1148	PP1150	>>>
16	Pit051	1349036	1349140	105	+	PP1173	PP1174	<><
17	Pit052	1349617	1349735	119	+	PP1174	PP1175	<><
18	Pit054	1388590	1388487	104	-	PP1209	PP1210	<<>
19	Pit055	1440293	1440130	164	-	PP1259	PP1260	<<>
20	Pit056	1441860	1442022	163	+	PP1259	PP1260	>><
21	RnpB/P28*	1512685	1513069	385	+	PP1326	PP1328	>>>

22	Pit058	1626951	1627100	150	+	PP1426	PP1427	<>>
23	T44*	1785122	1785242	121	+	PP1590	PP1591	<>>
24	RsmZ*	1822011	1822190	180	+	PP1624	PP1625	>><
25	Pit064	1847250	1847088	163	-	PP1652	PP1653	><>
26	RNA1**	1995669	1995866	198	+	PP1781	PP1782	<><
27	Pit077	2151206	2150992	215	-	PP1905	PP1906	<<>
28	RgsA/P16*	2229834	2229726	109	-	PP1967	PP1968	><>
29	C4 AS RNA 3*	2302915	2302823	93	-	PP2027	PP2026	<<<
30	Rmf*	2388735	2388345	391	-	PP2095	PP2096	><>
31	Pit092	2435418	2435212	207	-	PP2133	PP2134	><>
32	RNA2**	2608031	2608171	141	-	PP2284	PP2285	><>
33	Pit094	2622634	2623131	498	+	PP2294	PP2295	>>>
34	Pit097	2672559	2672471	89	-	PP2339	PP2340	<<>
35	Pit098	2674735	2674968	234	+	PP2343	PP2344	>><
36	RNA3**	2710973	2710798	176	-	PP2373	PP2374	><<
37	C4 AS RNA 6*	2855850	2855745	106	-	PP2507	PP2508	><>
38	Pit105	2925591	2925749	159	+	PP2563	PP2564	<><
39	Pit107	2939084	2939246	163	+	PP2570	PP2571	<><
40	Pit109	3261547	3261423	125	-	PP2859	PP2858	<<<
41	Pit110	3275580	3275861	282	+	PP2873	PP2874	>>>
42	Pit114	3450217	3450305	89	+	PP3067	PP3068	>>>
43	P15*	3466266	3466060	207	-	PP3080	PP3081	<<<
44	Pit124	3826437	3826208	230	-	PP3381	PP3380	<<<
45	Pit125	3828006	3828169	164	+	PP3381	PP3382	<><
46	Pit126	3967909	3967809	101	-	PP3497	PP3498	>>>
47	Pit127	3971957	3971765	193	-	PP3501	PP3502	><>
48	Ps2*	4013251	4013565	315	+	PP3540	PP3541	>><
49	Pit130	4022619	4022473	147	-	PP3548	PP3547	<<<
50	Pit132/IGR 3586	4073874	4073623	252	-	PP3586	PP3585	<<<
51	Pit133	4075444	4075605	162	+	PP3587	PP3586	<><
52	Pit136	4224280	4224507	228	+	PP3703	PP3704	<><
53	Pit137	4302495	4302290	206	-	PP3774	PP3775	><<

<b>54</b>	Pit140/IGR 3917	4425377	4425164	214	-	PP3916	PP3917	><>
<b>55</b>	PrrF2*	4595167	4595310	144	+	PP4069	PP4070	>>>
<b>56</b>	Pit144	4595281	4595233	49	-	PP4069	PP4070	><>
<b>57</b>	IGR 4095	4630733	4630507	227	-	PP4094	PP4095	><>
<b>58</b>	Bacteria small SRP*	4858503	4858392	112	-	PP4273	PP4274	><<
<b>59</b>	Pit148/IGR 4451	5047215	5047412	198	+	PP4450	PP4451	>><
<b>60</b>	Pit149	5103279	5103410	132	+	PP4491	PP4492	>>>
<b>61</b>	Pit151	5140624	5140398	227	-	PP4524	PP4525	><<
<b>62</b>	C4 AS RNA 4/IGR 4535*	5148997	5148876	122	-	PP4535	PP4534	<<<
<b>63</b>	Pit153	5219064	5218924	141	-	PP4598	PP4599	>>>
<b>64</b>	Pit154	5222758	5222598	161	-	PP4603	PP4602	<<<
<b>65</b>	Pit155	5224327	5224489	163	+	PP4603	PP4604	<><
<b>66</b>	PrrF1*	5325394	5325485	92	+	PP4685	PP4686	>><
<b>67</b>	CrcZ*	5338284	5338625	342	+	PP4696	PP4697	>>>
<b>68</b>	P31*	5373151	5373213	63	+	PP4724	PP4725	<><
<b>69</b>	P32*	5373351	5373255	97	-	PP4724	PP4725	<<<
<b>70</b>	SsrA tmRNA*	5389989	5390412	424	+	PP4738	PP4739	>>>
<b>71</b>	IGR 4740	5391608	5391314	295	-	PP4739	PP4740	><>
<b>72</b>	P24*	5437800	5437675	126	-	PP4775	PP4776	<<>
<b>73</b>	Pit161	5453307	5453144	164	-	PP4790	PP4791	><>
<b>74</b>	Pit162	5545496	5545288	209	-	PP4879	PP4878	<<<
<b>75</b>	Pit163	5756969	5756716	254	-	PP5049	PP5050	<<>
<b>76</b>	6S RNA/SsrS*	5934661	5934846	186	+	PP5202	PP5203	>>>
<b>77</b>	Pit168	5989892	5989792	101	-	PP5247	PP5248	><<
<b>78</b>	Pit169	6039010	6039211	202	+	PP5290	PP5291	<><
<b>79</b>	Pit172	6137172	6137302	131	+	PP5384	PP5385	>>>
<b>80</b>	Pit176	6159099	6158994	106	-	PP5401	PP5402	><<

\* Annotated sRNA.

\*\* Additional sRNAs not identified in Bojanovič *et al.*, manuscript in preparation.

**Table S7. Putative actuatons.**

For each transcript the name, start and stop coordinates, length, strand, 5' and 3' flanking genes and orientation relative to the flanking genes are reported.

Number	Name	Start	Stop	Length	Strand	5'Flanking gene	3'Flanking gene	Orientation
1	RNA4	611063	610868	195	-	PP0525	PP0526	<<>
2	RNA5	1532000	1531903	97	-	PP1344	PP1345	<<>
3	RNA6	1607721	1607616	105	-	PP1409	PP1408	<<<
4	RNA7	1748828	1748728	100	-	PP1549	PP1548	<<<
5	RNA8	2532043	2532142	99	+	PP2222	PP2223	>>>
6	Pit108	3023086	3023256	170	+	PP2638	PP2639	>>>
7	RNA09	4170053	4170158	105	+	PP3668	PP3669	<>>
8	RNA10	4564500	4564595	95	+	PP4049	PP4050	<>>

**Table S8. Putative ORFs.**

Putative ORFs are identified by the gene finders GLIMMER and GeneMark.

Table A. lists the ORFs that have been found in both gene finders and have the same translational coordinates. Table B. lists the ORFs that show a different translational start site between the two gene finders but same stop site.

For each putative ORF, coordinates, strand, length, predicted TSS, flanking genes, orientation and Blastp result are reported.

**A. Novel ORFs with same translational coordinates**

<b>ORF<sup>a</sup></b>	<b>Coordinates</b>	<b>Strand</b>	<b>Length (bp)</b>	<b>Predicted TSS</b>	<b>5' Flanking gene</b>	<b>3' Flanking gene</b>	<b>Orientation</b>	<b>Blastp</b>
<b>PP0284.1</b>	343442-342999	-	444	343871	PP0284	PPT03	><>	hypothetical protein
<b>PP0636.1</b>	744476-744916*	+	441	743427	PP0636	PP0637	<><	hypothetical protein (p)
<b>PP0651.2</b>	759860-760087*	+	228	759513	PP0651	PP0652	>><	hypothetical protein
<b>PP0651.3</b>	758879-758760	-	120	759671	PP0651	PP0651	>< <sup>b</sup> >	hypothetical protein
<b>PP1115.1</b>	1275412-1275113**	-	300	1275760	PP1115	PP1116	<<<	hypothetical protein
<b>PP1810.1</b>	2037877-2037602	-	276	2038030	PP1810	PP1811	><>	DUF 3077 superfamily
<b>PP1935.1</b>	2182187-	+	393	2181461	PP1935	PP1936	<><	hypothetical protein

<b>PP2874.1</b>	2182579* 3275654- 3275842	+	189	3275596	PP2874	PP2875	> > <	hypothetical protein
<b>PP3108.2</b>	3516293- 3516916**	+	624	3515633	PP3108	PP3109	> > >	rhs family protein
<b>PP3108.4</b>	3516127- 3516246	+	120	3515633	PP3108	PP3109	> > >	type IV secretion protein rhs-like protein
<b>PP3688.1</b>	4197678- 4197145*	-	534	4198048	PP3688	PP3689	< < >	hypothetical protein (P)
<b>PP4535.2</b>	5152241- 5151921*	-	321	5152397	PP4535	PP4536	< < >	hypothetical protein (P)

#### B. Novel ORFs with different start codon position

ORF <sup>a</sup>	Start site Glimmer/G eneMark	Stop site	St ra nd	Length Glimmer/G eneMark (bp)	Predicted TSS	5'Flanki ng gene	3'Flanki ng gene	Orientatio n	Blastp
<b>PP1115.2</b>	1276318/ 1276015	1276452	+	135/438	1274407	PP1115	PP1116	< > <	hypothetical protein
<b>PP1919.1</b>	2163137*/ 2163278	2164150	+	1014/873	2162727	PP1919	PP1920	> > <	hypothetical protein (p)
<b>PP1935.3</b>	2183476/ 2183269**	2183766	+	291/498	2181461	PP1935	PP1936	< > <	hypothetical protein
<b>PP1935.4</b>	2184249/ 2183940*	2184506	+	258/567	2181461	PP1935	PP1936	< > <	ser recombinase superfamily, HTH hin like

<b>PP1936.1</b>	2188357/ 2188522*	2187608	-	750/915	2188873	PP1936	PP1937	< < >	superfamily <b>(P)</b> hypothetical protein <b>(P)</b>
<b>PP2294.1</b>	2623027/ 2622829**	2623161	+	135/333	2621650	PP2294	PP2295	> > >	hypothetical protein
<b>PP2509.1</b>	2857541/ 2857313**	2857699	+	159/387	2856899	PP2509	PP2510	< > <	diadenosine tetraphosphate hydrolase <b>(P)</b>
<b>PP3066.3</b>	3448342/ 3448378	3447872	-	471/507	3450644	PP3066	PP3067	> < >	hypothetical protein <b>(P)</b>
<b>PP5237.1</b>	5971944/ 5971962	5971831	-	114/132	5975951	PP5238	PP5237	> < >	hypothetical protein (p)

<sup>a</sup> ORF name is assigned based on the 5' flanking gene and consecutive number from Frank *et al.*.

<sup>b</sup> Internal and on the opposite strand of the gene.

\* Found also in Frank *et al.* with the same coordinates.

\*\* Found also in Frank *et al.* with a different start position.

(p) Highly conserved in *Pseudomonas putida*.

**(P)** Highly conserved in *Pseudomonas spp.*

**Table S9. Strains, plasmids and oligonucleotides used in this work.**

Strain	Genotype	Ref.
<i>P. putida</i> KT2440	<i>rmo- mod+</i>	DSMZ
<i>E. coli</i> NEB5 $\alpha$	<i>fhuA2</i> $\Delta$ ( <i>argF-lacZ</i> )U169 <i>phoA</i> <i>glnV44</i> $\Phi$ 80 $\Delta$ ( <i>lacZ</i> )M15 <i>gyrA96</i> <i>recA1</i> <i>relA1</i> <i>endA1</i> <i>thi-1</i> <i>hsdR17</i>	NEB
Plasmid	Genotype	Ref.
pPCV31	<i>xylS</i> , <i>gfp</i> expressed fom Pm promoter in pSEVA, pBBR1 origin of replication, Gm <sup>R</sup>	Unpublished
pISO1	pPCV31 with TPP riboswitch upstream of <i>gfp</i>	This work
pISO2	pISO1 with RBS between TPP riboswitch and <i>gfp</i>	This work
Oligonucleotide	Sequence	Application
GSP1 PP0147	GGCGGCGCAGCAGATCATGT	5' RACE of citrate transporter
GSP2 PP0147	GCGGTGCCGACCGAGACTTTCA	5' RACE of citrate transporter
GSP1 PP4010	GCGGCTCGCCCAAGGCTCTG	5' RACE of <i>cspD</i>
GSP2 PP4010	GCAGCGGCCGCGGCATCTTT	5' RACE of <i>cspD</i>
GSP1 PP1623	CGGCCGGCAGGGTCACCCCT	5' RACE of <i>rpoS</i>
GSP2 PP1623	CCGCTTTCGTCGCCGCTCTT	5' RACE of <i>rpoS</i>
RNA_adapter	GCUGAUGGCGAUGAAUGAACACUGCGUUUGCUGGCUUUGAUGAAA	5' RACE RNA adapter
Adapter_primer	GCTGATGGCGATGAATGAACACTGC	5' RACE adapter primer
ITD1	AGCTTGUCCAGCAGGGTTGTCCAC	USER cloning: construction of pISO1
ITD2	ACAAGCUGATGGACAGGCTGCG	USER cloning: construction of pISO1
ITD3	ATGGTCAUGACTCCATTATTATTGTTTCTGTTGC	USER cloning: construction of pISO1
ITD4	ATGACCAUGCCTAGGCCGCGGCCGCGCGCATTTACCTGCTTGCTTGCTGACC	USER cloning: construction of



<b>ITD5</b>	ATCGCTUTTTCTTGTTGGTCATCACAGG	pISO1 USER cloning: construction of pISO1
<b>ITD6</b>	AAGCGAUAACCTCAGCATGAGTAAAGGAGAAGAACTTTTCACTGGAG	USER cloning: construction of pISO1
<b>ITD32</b>	ATCAACCUCAGCGCTGAGGCGATAGGAGGAATATACCATGAGTAAAGGAGAAG AACTTTTCACTGGAG	USER cloning: construction of pISO2
<b>ITD33</b>	AGGTTGAUCGCTTTTCTTGTTGGTCATC	USER cloning: construction of pISO2
<b>ITD17</b>	GCGGAGCTATCCAACGGCGG	plasmid sequencing
<b>ITD18</b>	GGACAGGGCCATCGCCAATTGG	plasmid sequencing
<b>ITD19</b>	GCTCGCGGCCATCGTCCACA	plasmid sequencing
<b>ITD20</b>	CCGCCAATTCGTCGCCCCATG	plasmid sequencing
<b>ITD21</b>	CAGTGGAGAGGGTGAAGGTGATGC	plasmid sequencing
<b>ITD22</b>	GGCGACTGCCCTGCTGCGTA	plasmid sequencing

## **Research Article 2**

**Global transcriptional responses to oxidative, osmotic, and imipenem stress conditions in *Pseudomonas putida***

K. Bojanovič, I. D'Arrigo, K. S. Long. (2016).  
Manuscript in preparation.



**Global transcriptional responses to oxidative, osmotic, and imipenem stress conditions in *Pseudomonas putida***

Klara Bojanović<sup>1</sup>, Isotta D'Arrigo<sup>1</sup>, and Katherine S. Long<sup>1\*</sup>

<sup>1</sup> The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Allé 6, DK-2970 Hørsholm, Denmark

\* corresponding author: Katherine S. Long<sup>1</sup>; phone: +45 45258024; fax: +45 45258001; e-mail: kalon@bio.dtu.dk

Keywords: *Pseudomonas putida* KT2440, membrane stress, transcriptomics, sRNA, differential expression, RNA-seq

Running title: Transcriptional responses to stress in *P. putida*

## Summary

Bacteria cope with and adapt to stress by modulating gene expression in response to specific environmental cues. In this study the transcriptional response of *Pseudomonas putida* KT2440 to oxidative, osmotic, and imipenem stress conditions at two time points is investigated via identification of differentially expressed mRNAs and sRNAs. A total of 440 small RNA transcripts are detected, where 10% correspond to previously annotated sRNAs, 40% are novel intergenic transcripts and 50% are antisense to annotated genes. Twenty-two pairs of complementary intergenic sRNA transcripts have been revealed and many of the antisense transcripts are found opposite rRNA and tRNA genes. Each stress elicits a unique transcriptional response as far as the extent and dynamics of the differentially expressed genes, where the most widespread changes are observed with osmotic stress after 60 minutes followed by oxidative stress after 7 minutes. The data show a higher fraction of differentially expressed sRNAs with greater than 5-fold changes compared with mRNAs. The work provides detailed insights into the mechanisms through which *P. putida* responds to different stress conditions and increases understanding of bacterial adaptation in natural and industrial settings.

## Introduction

Bacteria commonly encounter stressful conditions during growth in their natural environments and in industrial biotechnological applications such as the biobased production of chemicals. As the coordinated regulation of gene expression is necessary to adapt to changing environments, bacteria have evolved numerous mechanisms to control gene expression in response to specific environmental signals. These include the activation of regulators including the general stress sigma factor RpoS (Battesti *et al.*, 2011; Landini *et al.*, 2014) and extracytoplasmic function sigma factors (Mascher, 2013). In addition, a wealth of two-component regulatory systems couples the sensing of environmental stimuli by a membrane-bound histidine kinase with a corresponding response regulator that modulates expression of specific genes (Capra and Laub, 2012).

Another class of regulators are the small regulatory RNAs, a heterogeneous group of molecules that are often expressed under specific conditions and in response to stress (Waters and Storz, 2009; Gottesman and Storz, 2011; Storz *et al.*, 2011). Although some act by binding to protein targets and sequestering their function, the majority bind to mRNAs via base pairing and regulate their expression by modulating translation and/or stability. The base-pairing sRNAs are divided into two groups according to their genomic location relative to their target(s). The *cis*-encoded or antisense sRNAs are encoded just opposite of and have perfect complementarity with their targets. The *trans*-encoded sRNAs are encoded in a different location than their targets and typically exhibit limited complementarity with their targets. Thus, they often have multiple targets and are incorporated into larger regulatory networks. In some bacteria the RNA chaperone Hfq facilitates interactions between *trans*-encoded sRNAs and their targets.

*Pseudomonas putida* has served as a laboratory model organism for environmental bacteria and thrives in a variety of terrestrial and aquatic environments, including strains that colonize the rhizosphere and soil contaminated with chemical waste (Nikel *et al.*, 2014). Although some characteristics including a versatile metabolism and general robustness towards stresses are shared with other pseudomonads, *P. putida* lacks virulence factors and has superior tolerance to organic solvents. These traits and the availability of tools for genetic manipulation make it an attractive host for applications in industrial biotechnology and synthetic biology (Poblete-Castro *et al.*, 2012).

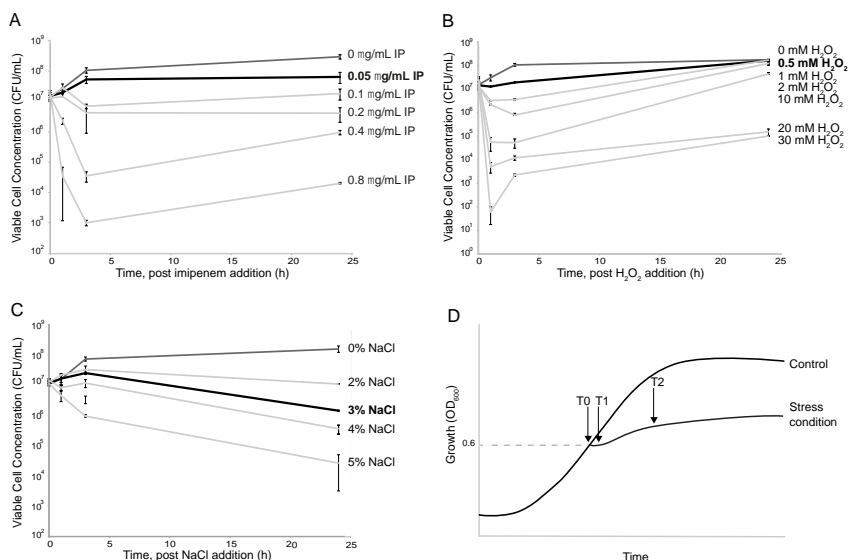
In this work, the complete transcriptional response (including mRNAs and sRNAs) of the well-characterized *P. putida* strain KT2440 to oxidative, osmotic and imipenem stress conditions is mapped with RNA-sequencing. A total of 440 small RNA transcripts are detected, consisting of both intergenic and antisense transcripts, where over half are conserved within the *Pseudomonadaceae* family. Each type of stress is found to elicit a unique pattern of transcriptional changes with respect to both the extent and dynamics of the response. In all stress types a general upregulation of genes encoding efflux pumps and other transporters, universal stress proteins and redox enzymes is observed. Specific alterations observed include an upregulation of beta-lactamase domain proteins under imipenem stress, induction of the SOS response and translational arrest under oxidative stress, and the accumulation of osmoprotectants and increased cardiolipin production under osmotic stress. The work identifies several small RNAs with differential expression in multiple stress conditions that are interesting targets for further functional characterization.

## Results

### *Experimental strategy*

In order to obtain a detailed understanding of *P. putida* stress response mechanisms, a RNA sequencing approach was used to investigate differentially expressed transcripts under oxidative, osmotic and membrane stress conditions. *P. putida* KT2440 was grown in minimal medium in the presence of hydrogen peroxide, sodium chloride, or the cell wall-targeting antibiotic imipenem to induce oxidative, osmotic, and membrane stress, respectively. With the aim of applying the maximal stress without affecting cell viability, a series of growth experiments were carried out with a range of different compound concentrations to determine the pseudo-steady state condition (Nielsen *et al.*, 2009), where there was nearly no change in growth or viability relative to that at compound addition. Growth and survival after compound addition were monitored via OD<sub>600</sub> and CFU counting, respectively (Fig. 1A-C). The final compound concentrations of 3 % NaCl, 0.05 mM H<sub>2</sub>O<sub>2</sub> and 0.1 µg/mL of imipenem were chosen to induce pseudo-steady state conditions. Cells were grown to mid-exponential phase, followed by compound addition and harvest after 7 minutes (T1) to investigate early transcriptional responses and 60 minutes (T2) of growth to observe the longer-term stress adaptation mechanisms. The control samples (T0) were harvested just prior to compound addition (Fig. 1D).





**Figure 1.** Effect of the addition of imipenem,  $H_2O_2$ , and NaCl on *P. putida* KT2440 survival as determined by viable cell concentration (CFU). Compounds were added to mid-exponential stage cultures in different concentrations, as marked on the right of each graph. The CFU count data after compound addition is shown. The chosen concentration of each compound is indicated in bold. (A) Effect of the addition of different concentrations of imipenem. (B) Effect of the addition of different concentrations of  $H_2O_2$ . (C) Effect of the addition of different concentrations of NaCl. (D) Representative growth curves for the chosen conditions. The stress experiments were performed by addition of the compounds in mid-exponential growth phase. Cells were harvested just before compound addition for the control (T0) and 7 minutes (T1) and 60 minutes (T2) after compound addition for the stress samples.

Following RNA isolation and library preparation, the samples were sequenced on the Illumina HiSeq platform. A total of 225 M reads were obtained, of which 200 M reads mapped to the *P. putida* KT2440

genome and on average 32% of the reads mapped to rRNA (Table S1). As 17-47% of the reads mapped to protein coding regions and 14-43% mapped to intergenic regions, the sequencing depth achieved in this study was very deep.

### *Identification of small RNA transcripts*

For small RNA identification, transcripts detected by Rockhopper were manually curated using Integrative Genomics Viewer (IGV) (Thorvaldsdottir *et al.*, 2013). One group of transcripts located in intergenic regions and having independent expression profiles relative to flanking genes was classified as novel intergenic sRNA candidates. A second group of transcripts encoded on the opposite strand relative to and having either partial or complete overlap with annotated genes was classified as *cis*-encoded antisense sRNAs. A total of 440 small RNA transcripts were identified in *P. putida* KT2440, significantly increasing the number of small RNA transcripts previously detected in this strain (Frank *et al.*, 2011). A total of 45 sRNAs were identified that are either annotated or have homology to known RNA motifs (Rfam) (Griffiths-Jones *et al.*, 2003) (Table S2). All sRNAs that are homologous in different *Pseudomonas* species (Gómez-Lozano *et al.*, 2015) were detected including: 6S RNA/SsrS, 4.5S RNA/Ffs/SRP, RNase P RNA, tmRNA/SsrA, PrrF1 and PrrF2, RsmY and RsmZ, CrcZ and CrcY, Spot42/ErsA/spf, t44, RgsA/P16, P26, and the FMN and TPP riboswitches. Seven copies of transcripts homologous to c4 antisense RNA were found (Citron and Schuster, 1990; Weinberg *et al.*, 2010). Three cobalamin riboswitches were detected in the KT2440 strain, as reported previously in the *P. putida* DOT-T1E strain (Molina-Santiago *et al.*, 2015), while four are known in *P. aeruginosa* (Gómez-Lozano *et al.*, 2012). Some annotated sRNAs including 6S/SsrS and t44 RNA were

manually identified as Rockhopper did not detect them despite very high expression profiles.

A total of 178 novel intergenic sRNA candidates were identified (Table S3) and denoted Pit001 to Pit178 for *Pseudomonas putida* intergenic transcript based on their genomic coordinates. The transcripts range in size from 24 to 1790 nt, with an average of 174 nt in length (Fig. 2A). Eight transcripts (Pit023, Pit053, Pit059, Pit062, Pit067, Pit098, Pit109, and Pit110) are putative 3'-UTR-derived sRNA candidates that overlap with the 3'-end of the gene or are in very close proximity of the stop codon (Chao *et al.*, 2012). Five transcripts (Pit014, Pit054, Pit057, Pit102, Pit108) are putative 5'-UTR-derived sRNA candidates or actuators (Kopf *et al.*, 2015).

A total of 217 *cis*-encoded RNA transcripts were identified (Table S4) and denoted Pat001 to Pat217 for *Pseudomonas putida* antisense transcript based on their genomic coordinates. These transcripts range in size from 21 to 1612 nt, with an average of 223 nt in length (Fig. 2A), which represent antisense transcription to 3.3% of the annotated genes in *P. putida* KT2440. In some cases, more than one antisense transcript is detected to the same gene. They are found to overlap the 3'-end, 5'-end, middle or even the entire gene on the opposite strand. The tRNA and rRNA genes had the largest number of antisense transcripts, followed by genes encoding hypothetical proteins (Fig. 2D). Many of the novel Pit and Pat RNA transcripts identified in this study have a Rho-independent terminator or a palindrome at the 3'-end (Table S3 and S4).

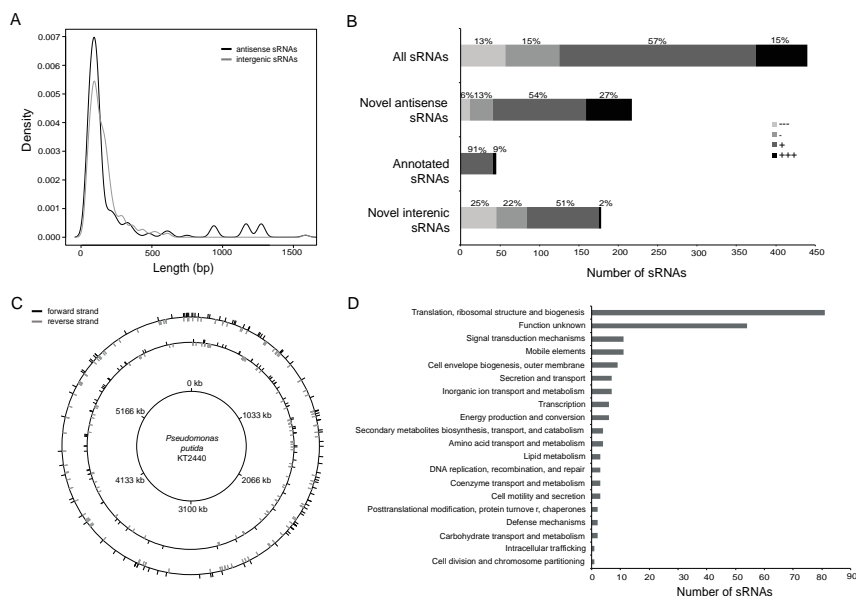
A total of 22 pairs of small RNA transcripts that are complementary in at least part of their sequences have been found (Table S5) and could potentially be acting as RNA sponges (Lalaouna *et al.*, 2015; Miyakoshi *et al.*, 2015). P30, an antisense transcript to CrcZ RNA has been detected in this study as reported previously in *P. aeruginosa* (Livny *et al.*, 2006;

Sonnleitner *et al.*, 2009). In addition to CrcZ having an antisense transcript, CrcY has two antisense transcripts, Pit118 and Pit119 in *P. putida* KT2440. Antisense transcripts were also detected to the tmRNA/SsrA (Pit157, Pit158), RsmZ (Pit063), RsmY (Pit020), 6S/SsrS (Pit164), P24 (Pat203), PrrF2 (Pit144), rmf RNA motif (Pit090) and SRP/4.5S RNA (Pit145). An antisense transcript to PrrF2 has been reported previously in *P. syringae* (Filiatrault *et al.*, 2010). In the case of 20 of these small RNA pairs, the two transcripts are encoded in the same genomic location just opposite each other, but in two cases the small RNA transcripts are encoded in distal genomic locations relative to each other (Pit146-Pit167 and Pit130-Pat180).

The novel sRNA transcripts found in this study were investigated for sequence conservation and homology in other bacteria using the Basic Local Alignment Search Tool (BLAST) (search criteria: query>80%, identity>60%, E<10<sup>-6</sup>) (Fig. 2B). For both the intergenic and antisense transcripts, approximately half are shared among bacteria in the *Pseudomonadaceae* family. Most of the other intergenic transcripts are only found either in the KT2440 strain or other *P. putida* strains, with only 2% being shared in other bacterial families. For the antisense transcripts, 19% are strain or species-specific, while 27% are shared among many bacterial families (Table S6). The latter is not surprising as a significant number of the antisense transcripts are located opposite various essential genes, including rRNA genes that are present in multiple copies (Table S4). Of all the 440 small RNAs identified in this study, 13% are strain specific, 15% are species specific, 57% are found among different bacteria in the *Pseudomonadaceae* family and 15% are found in other families.

The chromosomal positions of the novel sRNAs transcripts are illustrated in Fig. 2C. For both the intergenic (outer circle) and antisense

(inner circle) transcripts the localization is evenly distributed on the genome. In order to search for homology among the novel RNA transcripts, the small RNA sequences were compared with each other using BLASTN (search criteria: query>80%, identity>60%,  $E < 10^{-6}$ ). Twenty-one groups of homologous sRNAs were identified (Table S7), including the previously known examples PrrF1-PrrF2 and CrcX-CrcY-CrcZ (Sonnleitner and Haas, 2011). The majority of homologous intergenic sRNAs are related to transposases and the homologous *cis*-encoded sRNAs are antisense to rRNA, tRNA or transposase genes. These groups may be regarded as ‘sibling sRNAs’ that can either be functionally redundant or exert non-redundant regulatory functions (Caswell *et al.*, 2014).



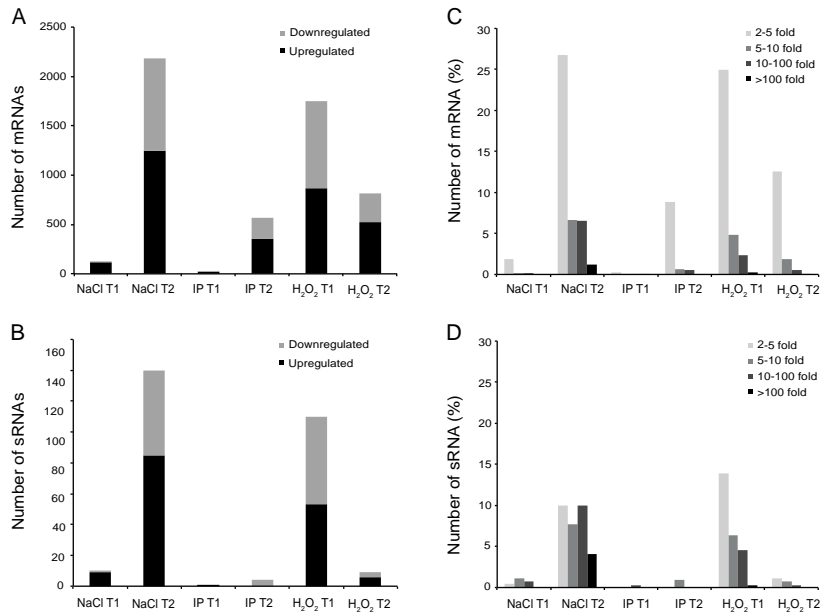
**Figure 2.** Properties of the small RNA transcripts identified in *P. putida* KT2440. (A) Length distribution of intergenic and antisense sRNA candidates. (B) Conservation of novel sRNA candidates: (---) no sequence conservation found outside of *P. putida* KT2440 strain; (-) no sequence conservation found outside of *P. putida* species; (+) sequence conservation primarily in *Pseudomonadaceae*; (+++) sequence conserved in bacterial species outside the *Pseudomonadaceae* family. (C) Genomic distribution of intergenic sRNAs (outside circle) and antisense sRNAs (inside circle), where the sRNAs encoded on the positive and negative strands are indicated on the outside and inside of the circles, respectively. (D) The numbers of *cis*-encoded sRNA candidates encoded opposite different functional classes of annotated genes.

*General patterns of mRNA and sRNA differential expression under stress conditions*

Each induced stress yields a particular transcriptional response in terms of the extent and dynamics of the gene expression changes. In terms of the number of differentially expressed transcripts in the different stress conditions, the mRNA and small RNA transcripts followed a similar pattern. The highest number of differentially expressed mRNAs and sRNAs (fold change  $\geq 2$ , p-value  $\leq 0.05$ ) compared to the control was observed with osmotic stress after 60 minutes, followed by oxidative stress after 7 minutes (Fig. 3A-B). Osmotic stress 7 minutes after the addition of NaCl had around 2% of transcripts changed in both groups of RNA. The effect in T2 had changed levels of transcription for almost 32% of sRNA and 41% of mRNA. On the other hand the oxidative stress caused the biggest changes in expression in short term (T1) with changes of 25% and 32% of sRNA and mRNA, respectively. These changes were less extensive in T2 and decreased to 15% for mRNA and 2% for sRNA. Membrane stress initiated with the antibiotic imipenem does not cause as severe adjustments as the other conditions tested (changes of mRNA and sRNA are under 1%). Only an hour after the addition of the antibiotic 10% of mRNA levels are diversified.

The observed fold-changes in expression under different stress conditions for mRNA and sRNA transcripts are summarized in Fig. 3C-D. The majority of mRNA transcripts exhibited 2-5 fold expression changes in all conditions and a higher proportion of mRNAs showed changes in this range compared to sRNAs. There was a higher fraction of sRNAs with above 5-fold expression changes compared to mRNAs in all stress conditions. Very high changes (above 100-fold) were observed for 4% of sRNA and 1% of mRNA transcripts during osmotic stress after 60 minutes. Although the fold-changes in sRNA and mRNA expression are

different, the patterns of the changes are similar. Osmotic and membrane stresses induced expression changes that increased with time while oxidative stress induced a strong immediate response that decreased after one hour.



**Figure 3.** An overview of the differentially expressed mRNAs and sRNAs. The number of differentially expressed mRNAs (A) and sRNAs (B) in osmotic (NaCl), imipenem (IP) and oxidative (H<sub>2</sub>O<sub>2</sub>) stress conditions at T1 (7 minutes) and T2 (60 minutes) compared to the control (without added stressor) are shown. The percentages of transcripts exhibiting different fold-changes in expression for mRNA (C) and sRNA (D) relative to the total number of 5350 CDS and 440 sRNAs, respectively.



### *Differential expression of mRNAs under osmotic stress*

The RNA expression profile of *P. putida* KT2440 exposed to NaCl revealed a much stronger response at 60 minutes compared to 7 minutes after NaCl addition, with 2182 (40.8% coding sequences CDS) and 124 (2.3% CDS) differentially expressed genes, respectively (Fig. 3A, Table S8). Only 80 genes (3.8% CDS) are common to both time points, where more than half encode proteins with an unknown function and several are transcriptional regulators (Fig. 4A). Out of the 124 differentially expressed genes at T1, only 18 had greater than 5-fold changes in expression (15 upregulated and 3 downregulated) and 12 are hypothetical proteins. At T2 there were 748 genes with greater than 5-fold changes in expression (627 upregulated and 121 downregulated).

The large number of differentially expressed genes observed under osmotic stress is due in part to the differential expression of many sigma factors and transcriptional regulators, suggesting that many regulatory networks were affected. Among the upregulated regulator genes are *rpoH*/ $\sigma^{32}$  (6-fold) that controls expression of heat shock proteins, the  $\sigma^E$  regulatory protein *mucB/rseB* and the anti-sigma factor *rseA* associated with biofilm formation (28- and 21-fold, respectively), the sigma factor ( $\sigma^{24}$  homolog) PP\_3577 (6-fold) and anti-sigma factor *fecR* (26-fold) connected to iron uptake. The sigma factor  $\sigma^{22}/algU$ , a repressor of flagella genes was also upregulated (17-fold) and sigma factor *sigX* that is involved in outer membrane composition was upregulated 57-fold. On the contrary the house-keeping  $\sigma^{70}$  was dramatically downregulated (42-fold). There were also many downregulated genes encoding regulators from different families (AsnC, Fis, ArsR, GentR, MarR, Cro/C1, AraC, TetR).

For the non-specific response to osmotic stress, the chaperones *groES*, *dnaK* and *dnaJ* (310-, 7-, 9-fold, respectively), heat shock proteins *hsp20*

and *hsp90* (13- and 10-fold, respectively) and two universal stress proteins (PP\_3237, PP\_2187) were upregulated (around 8-fold). In addition *recA* and the catalases *katA* and *katE*, involved in the general ROS response, were upregulated (3-, 3- and 281-fold, respectively). Interestingly, the cold shock protein *cspA* was downregulated 6-fold. Moreover, 12 genes related to biofilm formation were >5-fold upregulated and flagella genes were downregulated (between -16 and -47-fold).

Specific responses to osmotic stress include two main mechanisms: (1) accumulation by import (proline and glycine betaine) or biosynthesis of osmoprotectants (trehalose, mannitol, N-acetylglutaminylglutamine amide (NAGGN)) in order to restore the osmotic balance in cells and (2) altering membrane composition by fatty acid saturation and phospholipid composition to endure the changed turgor pressure (Brown, 1976; Romantsov *et al.*, 2009). The osmoprotectant operon *opuBC-BB-BA* for glycine/proline betaine uptake, the proline betaine MFS transporter *proP*, and two members of the choline/carnitine/betaine transporter family were highly upregulated (above 5-fold). In *P. putida* KT2440 the trehalose synthesis pathway was upregulated, and showed a very close similarity to *P. aeruginosa* (Djonović *et al.*, 2013) in terms of operon composition and biosynthetic pathway. The PP\_4051-4054 (predicted *treZY*) and PP\_4058-4059 (predicted *treS*) operons, the single-gene PP\_4060 (alpha-amylase) and the glycogen metabolism genes PP\_2918 and PP\_4050 were highly expressed in osmotic stress. The two genes PP\_1748 and PP\_1750 show similarity to *P. aeruginosa* NAGGN biosynthetic genes (Aspedon *et al.*, 2006) and have very high upregulation at T2. Moreover, mannose synthesis was activated, with phosphomannomutase (PP\_5288) and *algA* (PP\_1277) genes upregulated 5- and 3-fold, respectively.

Genes for cardiolipin production involved in membrane alteration are usually responsive to osmotic stress. The cardiolipin synthase 2 (PP\_3264) and the entire predicted operon (PP\_3263-3266) were strongly upregulated. Part of the cell defense towards osmotic stress is to change the transcription of different transporters. Upregulation of some RND efflux pumps (operon PP\_5173-5175, PP\_3302-3304, *ttg2* operon), permeases, ABC transporters was observed as well as downregulation of several other transporter-related proteins (21 were downregulated over 5-fold).

#### *Differential expression of mRNAs under oxidative stress*

The RNA expression profile of *P. putida* KT2440 exposed to hydrogen peroxide revealed a much stronger response at 7 minutes compared to 60 minutes after compound addition, with 1746 (32.6% CDS) and 814 (15.2% CDS) differentially expressed genes at T1 and T2, respectively (Fig. 3A, Table S9). Almost one-fifth (409) of the differentially genes at T1 also have changed transcriptional levels at T2 (Fig. 4B). Substantial changes in mRNA levels (fold  $\geq 5$ ) were observed for 382 and 129 genes at T1 and T2, respectively. Bacteria typically encode several enzymes that help in detoxification of ROS, including catalases and peroxiredoxins. The data shows that the major catalase gene *kata* (PP\_0481) was upregulated more than 900-fold at T1 and more than 20-fold at T2, while *katB* (PP\_3668) was upregulated more than 30-fold at T1 and almost 6-fold at T2. Also the expression of the two hydroperoxide reductases *ahpC* (PP\_2439) and *ahpF* (PP\_2440) was very high at T1 (247- and 334-fold) and then decreased at T2 (2- and 4-fold). Genes *kata*, *katB*, *ahpC* and two thioredoxin reductases (*trxB*, *trx-2*) are under the OxyR redox-sensing regulator (Yeom *et al.*, 2012),

which is constitutively expressed and its regulon is activated in the presence of H<sub>2</sub>O<sub>2</sub> (Kim and Park, 2014) as shown in our data in T1.

Two regulators have been previously shown to be important in the maintaining intracellular redox status: FinR (PP\_1637), which is essential for the induction of *fpr* (ferredoxin-NADP<sup>+</sup> reductases) during exposure to superoxide stress and HexR, which regulates intracellular energy and redox status by the Entner–Doudoroff pathway (ED) (Kim and Park, 2014). In our study none of these regulators neither its regulated genes had changed transcriptional levels. On the other hand upregulation in transcript levels of several other redox enzymes (cytochrome and quinone carrier proteins) has been observed.

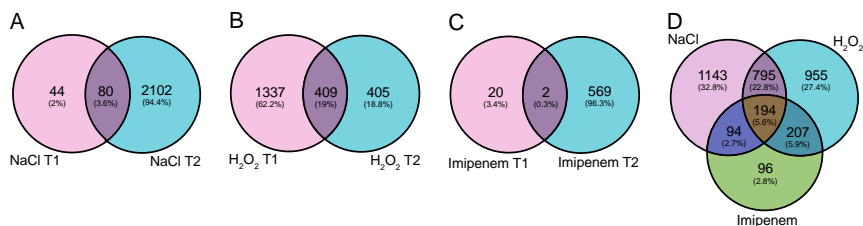
The present data confirmed the upregulation and downregulation of specific genes in a previous microarray study on the KT2440 strain under paraquat and cumen hydroperoxide stress (Yeom *et al.*, 2012). Many ribosomal proteins were downregulated, whereas several membrane proteins, transporters, and DNA repair mechanisms being upregulated. Interestingly, taurine transport and metabolism was upregulated in T1, consistent with the role of taurine as an antioxidant and membrane stabilizer.

#### *Differential expression of mRNAs under imipenem stress*

The transcriptome analysis showed that *P. putida* KT2440 exposed to the beta-lactam antibiotic imipenem had less drastic changes on the transcriptional level compared to the other two stress conditions. A total of 593 genes (Fig. 3A, Table S10) were differentially expressed; 22 (0.4% CDS) at T1 and 571 (10.7% CDS) at T2 (Fig. 4C). There was an overlap of only two common differentially expressed genes in both time points, encoding a hypothetical protein and a predicted membrane protein.

The genes with the highest fold changes at T1 are related to changed transcriptional levels of some membrane proteins including ABC and other transporters. At T2 43 genes are upregulated and 12 are downregulated with above 5-fold changes. Interestingly, a cluster of genes PP\_2663-PP\_2682 is upregulated more than 5-fold, which is the same cluster found to be upregulated in this strain exposed to chloramphenicol (Fernández *et al.*, 2012). This region includes a redox sensing protein, the AgmR regulator, an ABC efflux pump (regulated by AgmR), several redox-related proteins (quinoproteins and pyrroloquinoline quinone biosynthesis protein) and a beta-lactamase domain-containing protein (PP\_2676). Another highly upregulated region (PP\_0375-0380) in this study and the chloramphenicol study is related to the *pqq* genes involved in coenzyme PQQ biosynthesis that are also regulated by AgmR. Some changes were observed in genes related to the electron transfer chain (azurin, cytochrome c oxidase, and glycolate oxidase) and response to the DNA damage (pyocins). In addition, the housekeeping sigma factor  $\sigma^{70}$  was downregulated 6-fold at T2.

Fig. 4D shows the numbers of differentially expressed genes that are unique to a specific type of stress condition and common to two or three types of stress conditions. Osmotic and oxidative stress conditions have the highest number of common differentially expressed genes (795 genes). There are 194 differentially expressed genes that are common in all three types of stress conditions tested, and likely represent the general response of *P. putida* KT2440 to stress.



**Figure 4.** Venn diagrams illustrating the number of differentially expressed genes under (A) osmotic stress (NaCl), (B) oxidative stress (H<sub>2</sub>O<sub>2</sub>), (C) imipenem (IP) stress and (D) in all three stress conditions. The proportions of differentially expressed genes in a certain type of stress condition are shown in parentheses.

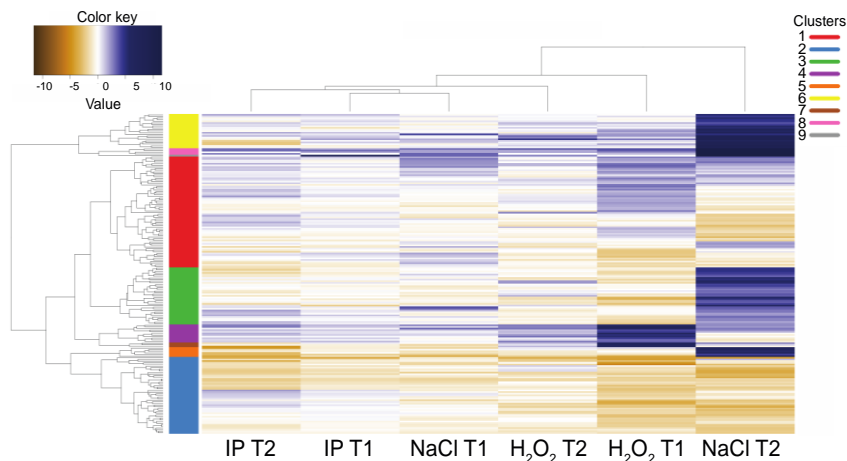
#### *Differential expression of small RNAs*

The largest numbers of differentially expressed small RNAs were observed in osmotic stress after 60 minutes followed by oxidative stress after 7 minutes. Taken together, 198 out of 440 sRNAs identified in this study were differentially expressed in at least one condition (Table 1; Table S11).

**Table 1.** Fold changes of eleven small RNAs with the differential expression in at least three out of six chosen conditions. IP stands for imipenem. Empty boxes show no differential expression in that condition. All sRNAs with differential expression are shown in Table S11.

Name	NaCl T1	NaCl T2	H <sub>2</sub> O <sub>2</sub> T1	H <sub>2</sub> O <sub>2</sub> T2	IP T1	IP T2
<b>Pat107</b>	-4.2	-13.5	-3.5	-3.5		-4.7
<b>Pat044</b>	8.7	7	71.5	7.6		
<b>Pat077</b>		-3.5	-2.9			-3.8
<b>Pit020</b>		-3.6	-3.8			-4.8
<b>RsmY</b>		-3.1	-3.7			-4.9
<b>Pat110</b>	6.8	6.1	4.2			
<b>Pit116</b>	5.5	5.8	4			
<b>Pit087</b>	5	8.1	2.9			
<b>Pat181</b>	4.8	4.7	7.6			
<b>Pit082</b>		-5.2	-3	-3.9		
<b>Pit080</b>		-12.8	-5.6	-4		

The differentially expressed sRNAs are clustered into nine groups (Fig. 5) based on their expression patterns in the different conditions.

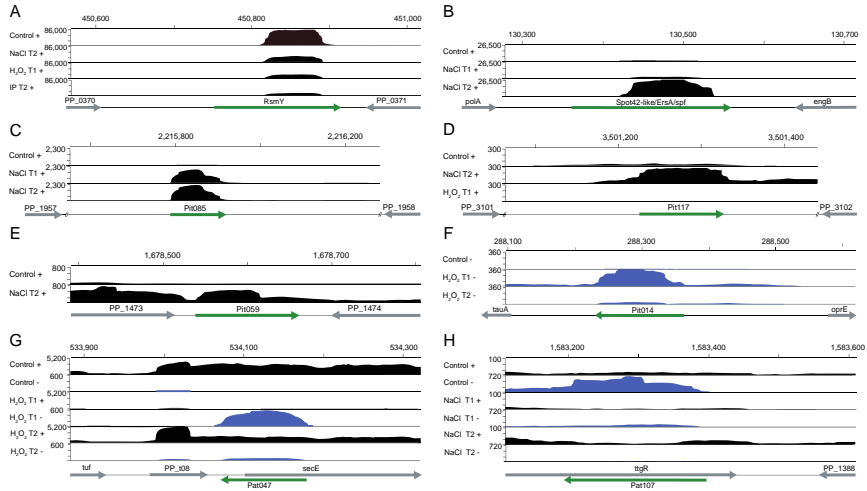


**Figure 5.** Heat map and hierarchical clustering of differentially expressed sRNAs in osmotic, oxidative and membrane stress conditions at T1 (7 minutes) and T2 (60 minutes) after exposure compared to the control without added stressor (fold change  $\geq 2$  and a  $p$ -value  $\leq 0.05$ ).

Three groups of sRNAs exhibit different extents of upregulation in osmotic stress after 60 minutes. Cluster 8 consists of four sRNAs with exceptionally high levels of upregulation (greater than 2000-fold), cluster 6 consists of sRNAs with 100-2000 fold changes, and cluster 3 includes transcripts with less than 100-fold changes. Clusters 4 and 7 consist of sRNAs highly expressed in oxidative stress T1, with some transcripts also being upregulated in other conditions (Table S11). The transcripts that are downregulated in all conditions group together in cluster 2. Pat092 comprises its own cluster (nr. 9) with high regulation in osmotic stress T2 and imipenem stress T1. The other two clusters (1 and 5) have sRNAs with either up- or downregulation in different conditions, showing a different regulation dependent upon the condition.



The expression profiles of selected annotated and novel sRNA transcripts exhibiting differential expression patterns are shown in Fig. 6.



**Figure 6.** Expression profiles of sRNAs in different conditions. The profiles include two annotated sRNAs RsmY (A) and Spot42/ErsA/spf (B), two novel intergenic sRNA candidates Pit085 (C) and Pit117 (D), a putative 3'UTR-derived sRNA candidate Pit059 (E), a putative 5'UTR-derived sRNA candidate or actuator Pit014 (F), and two novel *cis*-encoded sRNA candidates Pat047 (G) and Pat107 (H). Reads on the forward (+) and reverse (-) strands are denoted in black and blue, respectively. Note that the scales for the + and - strands differ. The sRNA transcripts are shown in green and the flanking genes are in gray. The genomic location is shown on the top.

The expression profiles of the two sRNAs RsmY and ErsA are shown in Fig. 6A-B. The ends of these transcripts are not visible as the central portion of the transcripts had a higher number of reads. The profiles of

four novel intergenic RNA transcripts are shown in Fig. 6C-D-E-F and two novel antisense RNAs are shown in Fig. 6G and 6H.

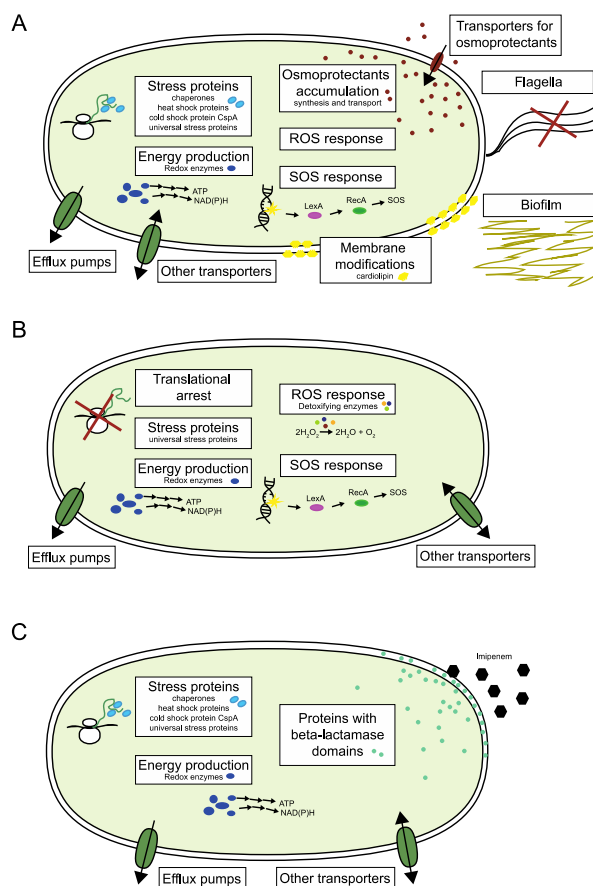
Only Pat107 sRNA (Fig. 6H) was differentially expressed and down-regulated in five out of six conditions. This sRNA is encoded opposite the *ttgR* gene (PP\_1387), which is a transcriptional repressor that regulates biosynthesis of antibiotics, efflux pumps, and osmotic stress (Ramos *et al.*, 2005). This gene was upregulated 3.1-fold in osmotic stress T2, where the highest down-expression for the *cis*-encoded sRNA Pat107 was observed (13.5-fold). The sRNA Pat077 was differentially expressed in three conditions and encoded opposite the *hexR* gene (PP\_1021), also a transcriptional regulator that is responsive to oxidative stress, but its levels were unchanged. RsmY (Fig. 6A) and Pit020 sRNA, which are antisense to each other were both 4-fold down-regulated in three conditions.

The differentially expressed *cis*-encoded sRNAs and their respective genes on the opposite strand form four groups regarding their expression patterns in the same condition. There are cases where a gene and its antisense RNA have fold-changes in the same direction (either both are up- or downregulated), where the two transcripts have opposite directions of expression; where the asRNA has different regulation in different conditions and the gene on the opposite strand has mixed directions; and cases where asRNAs are differently expressed but the genes are not. These patterns can provide clues as to the mechanism of asRNA regulation of gene expression (Thomason and Storz, 2010).

## Discussion

Our results show that *P. putida* KT2440 combines different defense mechanisms upon exposure to stressful conditions. Osmotic and membrane stress induce transcriptional changes that increase with time

while oxidative stress triggers rapid expression changes in order to survive. Extensive transcriptional changes suggest that the stress conditions studied here provoked specific and nonspecific responses. The first stage of a changed environment demands a rapid adjustment for survival followed by a second stage including alternations in metabolism and adaptive processes fitting the new conditions (Brown, 1976). Most relevant stress-induced mechanisms found to be responsive in *P. putida* KT2440 in this study are summarized in the Fig. 7.



**Figure 7.** Overview of the main mechanisms used by *Pseudomonas putida* KT2440 against (A) osmotic, (B) oxidative and (C) membrane stress caused by imipenem.

Many previously unnoticed genes from microarray studies were detected in the RNA-seq data in this study and found to be differentially expressed. This could be due to the higher sensitivity of RNA-seq as far as detecting lower abundance transcripts (Fu *et al.*, 2009; Zhao *et al.*,

2014). Many genes whose products are necessary for the defense mechanisms do not have a change in transcription levels because they are regulated on other levels are not caught by RNA-seq. A previous microarray study in *P. aeruginosa* exposed to hydrogen peroxide after 10 minutes detected 33,7% differential expression (Palma *et al.*, 2004), which concurs with this study with changes in the transcriptional levels of 32.6% of the CDS and a similar study in *E. coli* (Zheng *et al.*, 2001). When *P. putida* has been subjected to paraquat and cumen hydrogenperoxide (organic peroxides) only 1.7% and 2.1% of genes with differential expression were detected in a microarray study (Yeom *et al.*, 2012), showing that addition of inorganic H<sub>2</sub>O<sub>2</sub> causes much more drastic adjustment changes. When *P. aeruginosa* was subjected to osmotic stress, a microarray study found only 2.4% of genes being differentially expressed over 3-fold, while we observed 40.8% CDS changed, but they also used a much lower concentration of the salt (Aspedon *et al.*, 2006). Imipenem conditions have caused the least differential expression, and were the most similar to the control, as previously observed for antibiotics ampicillin, chloramphenicol and kanamycin (Molina-Santiago *et al.*, 2015). The extent of the differential expression observed in different studies is also due to the degree of stress applied.

A general bacterial response to osmotic stress is the accumulation of osmoprotectants to establish osmotic balance intracellularly and also protect enzyme activities. In different microorganisms different compatible solutes are found such as electrolytes (KCl), amino acids and their derivatives (proline, glycine, betaine, glutamine, aspartate), polyols (glycerol, arabitol), and sugars (trehalose, sucrose) (Brown, 1976; Galinski and Trüper, 1994). The present data show that accumulation of glycine/proline betaine by import uptake system, and biosynthesis of

NAGGN, trehalose, mannitol, and glycogen are important strategies for *P. putida* KT2440 to respond to osmotic stress. NAGGN, mannitol and trehalose have been shown to be important compatible solutes in pseudomonads previously (Kets *et al.*, 1996; Castro *et al.*, 2000; Aspedon *et al.*, 2006). NAGGN seems to be the most important osmoprotectant as genes for its biosynthesis were among the most upregulated genes in T2 and similar has been observed previously for *P. aeruginosa* (Aspedon *et al.*, 2006). Import of potassium ions is also a very common way to fight with osmotic stress in other organisms (Galinski and Trüper, 1994; Weber and Jung, 2002) but upregulation of these transporters was not observed here. Finally, upregulation of iron-uptake mechanisms (siderophores) have been observed here (15-46 fold), as it has been previously reported for *Sinorhizobium meliloti* (Rüberg *et al.*, 2003).

As second defensive mechanism to osmotic stress, bacteria change membrane composition by increasing cardiolipin content. Upregulation of the biosynthetic genes for cardiolipids has been previously observed in *B. subtilis* and *E. coli* (Romantsov *et al.*, 2009). This mechanism was also confirmed in *P. putida* as the expression of cardiolipin biosynthesis genes was highly upregulated.

Finally, a downregulation of flagellar genes and an upregulation of biofilm formation was observed in *P. putida*, which has been also previously observed in response to salt-stressed cells (Rachid *et al.*, 2000; Weber and Jung, 2002; Rüberg *et al.*, 2003; Mukhopadhyay *et al.*, 2006; Tsuzuki *et al.*, 2011). Motility reduction and biofilm formation seem to be a general bacterial response to osmotic stress.

*P. putida* has developed different mechanisms of oxidative stress sensing, regulation, and defense (reviewed in Kim and Park, 2014), among which upregulation of the detoxifying enzymes seems to be the

most drastic change in the presence of hydrogen peroxide. Their expression is controlled by several regulators, such as OxyR, FinR and HexR, involved in protection against ROS. The two major oxidative stress regulators in *E. coli* and *S. typhimurium* are SoxR and OxyR (Farr and Kogoma, 1991). However, in *P. putida* SoxR regulator is not responsive to oxidative stress (Yeom *et al.*, 2012) and the oxidative stress defense-genes of the SoxR regulon in enteric bacteria such as *fpr*, *fumC-1*, *sodA*, and *zwf-1* are independent of SoxR in *P. putida* (Park *et al.*, 2006). Moreover, these genes have been previously shown to be responsive to superoxide and nitric oxide in *P. putida* (Park *et al.*, 2006) but not in the presence of cumen hydroperoxide (Yeom *et al.*, 2012). Similarly to cumen hydroperoxide stress *fpr*, *fumC-1*, *sodA* and *zwf-1* were not differentially expressed in this study after adding H<sub>2</sub>O<sub>2</sub>. This suggests that they are specifically induced dependent upon the compound causing the oxidative stress.

OxyR is a transcriptional activator, which is constitutively expressed and activated by H<sub>2</sub>O<sub>2</sub> (Kim and Park, 2014). In this study its transcription levels did not change, while its responsive genes (*katA*, *katB*, *aphC*, *trxB*, *trx-2*, *hslO*) changed the expression. Our data shows that catalases and hydroperoxide reductases change transcriptional levels in the presence of hydrogen peroxide very fast and very drastically, since their expression was very soon after the induction of the oxidative stress and was lower but still significant 60 minutes after. Hydroperoxide reductase AphC has been shown not to be adequate to detoxify high levels of peroxide (Poole and Ellis, 1996), while the catalases have the major role in survival under its presence (Hishinuma *et al.*, 2006; Bignucolo *et al.*, 2013).

A part of oxidative stress defense is also redox-sensing regulators. *P. putida* KT2440 harbours FinR, and HexR, regulating *fpr* and *zwf* genes,

respectively. *fpr* gene encodes for the ferredoxin-NADP<sup>+</sup> reductase, which has been highly induced under paraquat in *P. putida* KT2440 (Lee *et al.*, 2006). We did not observe *fpr* gene induction in our data. The gene *zwf* also encoding the ferredoxin-NADP<sup>+</sup> reductase, has been upregulated in response to superoxide or nitric oxide (Park *et al.*, 2006) but transcriptional levels changes have not been observed here under hydrogen peroxide. Therefore, different compounds causing oxidative stress (Kim and Park, 2014), seem to activate different mechanisms and responses.

It has been previously shown that ED pathway, which generates redox molecules NAD(P)H is essential in *P. putida* for a higher tolerance to osmotic stress (Chavarría *et al.*, 2013). Ferredoxin-NADP<sup>+</sup> reductases also other enzymes in the ED pathway have not been upregulated in the presence to H<sub>2</sub>O<sub>2</sub> in this study, as downregulation of primary metabolism has been observed in *P. aeruginosa* (Palma *et al.*, 2004). These genes could be regulated on post-transcriptional level or show a general slowdown of metabolism in stress. The electron transport chain has been shown to be downregulated on the protein level (Bignucolo *et al.*, 2013) but we have detected upregulation in transcript levels of several redox enzymes (cytochrome and quinone carrier proteins). This is showing that bacteria need additional sources of energy to cope with oxidative stress. At the same time various ribosomal proteins have been downregulated as shown previously (Yeom *et al.*, 2012) and DNA repair machinery was upregulated. Taken together cells have made enormous rearrangements and control very carefully which proteins are necessary in such hostile environment, to not have any energy losses.

The upregulation of several SOS response genes (*lexA*, *recA*, and *recN*) was detected here at both time points during oxidative stress and after 60 minutes with osmotic stress. The SOS regulon is probably upregulated



indirectly by H<sub>2</sub>O<sub>2</sub> and NaCl by oxidant-induced DNA damage and a prolonged osmotic stress exposure. Similar changes have been observed in *P. aeruginosa* (Palma *et al.*, 2004) and *E. coli* (Zheng *et al.*, 2001).

Antibiotics also cause oxidative stress in the cells by increasing the levels of ROS, which inactivate various cell enzymes (Albesa *et al.*, 2004; Dwyer *et al.*, 2009; Kim and Park, 2014). Microarray study in *P. putida* and *P. aeruginosa* showed that the ampicillin (beta-lactame antibiotic) activated oxidative-stress and SOS inducible genes (Yeom *et al.*, 2010). Interestingly, we found the same region of genes (PP\_2663-2682) upregulated in cells exposed to imipenem as previously shown response to exposure to chloramphenicol (Fernández *et al.*, 2012), although these two antibiotics have different mechanisms of action. Surprisingly, the same region has been upregulated in cells exposed to H<sub>2</sub>O<sub>2</sub> in T2 (14-116-fold), while some of these genes are downregulated in osmotic stress (4-44-fold). Upregulation of the PP\_2669 gene has also been observed in the rhizosphere due to oxidative stress caused by antimicrobials in the environment (Matilla *et al.*, 2007), where the *pqq* genes are a part of the cell defense to redox changes/antioxidants (Misra *et al.*, 2004). This region seems to be important in the response to oxidative stress and antimicrobials causing oxidative stress.

The beta-lactamase genes *ampC*, *ampG*, and *ampD* were not upregulated in the presence of the imipenem in this study. This may be a longer exposure time is need to activate more pronounced changes in this specific response (Bagge *et al.*, 2004). On the other hand a beta-lactamase domain-containing protein (PP\_2676) was upregulated 60 minutes after imipenem addition, showing that the degradation of the antimicrobials is an important strategy.

Extrusion of molecules causing stress has previously been shown to be an important response for *P. putida* survival (Fernández *et al.*, 2012;

Molina-Santiago *et al.*, 2014, 2015; Ramos *et al.*, 2015). Indeed, changed transcriptional levels were detected in several permeases, ABC and RND efflux pumps in all chosen conditions. Downregulation of some transporters and upregulation of others indicates that the cells are very selective in which molecules are being transported via membrane in order to help the cell survival.

A larger fraction of sRNAs with higher fold-changes in expression compared to mRNAs was observed in all conditions. The reason could be because they exert regulatory roles and for that they sometimes have to compete with other regulators in order to change the ratios in the cell. In *P. putida* KT2440 previously 36 intergenic sRNAs have been detected out of which 22 are annotated sRNAs with homology in other *Pseudomonas* species (Frank *et al.*, 2011). In *P. putida* DOT-T1E strain 154 *trans*-encoded sRNAs have been found in a RNA-seq study (Molina-Santiago *et al.*, 2015). There has been a gap in the number of transcripts observed in a reference strain *P. putida* KT2440 comparing to other pseudomonads (Gómez-Lozano *et al.*, 2015). In *P. aeruginosa* 573 and 233 intergenic sRNAs have been reported in PAO1 and PA14, respectively with and overlap of 126 sRNAs in both strains (Ferrara *et al.*, 2012; Gómez-Lozano *et al.*, 2012; Wurtzel *et al.*, 2012).

The only functional characterization of sRNAs in *P. putida* thus far has been done with CrcZ/CrcY sRNAs, showing that it binds and titrates Hfq and prevents it from repressing target mRNAs in *P. aeruginosa* PAO1 (Sonnleitner and Bläsi, 2014). For the rest of the sRNAs only a homology to known motifs does not necessarily mean that they carry the same function in this strain. The majority of characterized sRNAs have been shown to have a function in pathogenic *P. aeruginosa* and are connected to its virulence, while *P. putida* KT2440 is avirulent (Nelson *et al.*, 2002). For example, sRNA PhrS is an activator of PqsR synthesis,

one of the key quorum-sensing regulators in *P. aeruginosa* but the PqsR protein is found only in *P. aeruginosa* strains (Sonnleitner *et al.*, 2011). In the present data PhrS is 3.3-fold downregulated in oxidative stress T1. PhrS sRNA must have different targets in other strains and possibly also in *P. aeruginosa*.

RsmY and RsmZ are two redundant sRNAs, which control secondary metabolism in pseudomonads by sequestering RsmA protein (Sonnleitner and Haas, 2011). Only RsmY was downregulated 3-5 fold in osmotic stress T2, oxidative stress T1 and membrane stress T2. Their activator *gacA* (PP\_4099) is also not differentially expressed in any of the chosen conditions and therefore probably does not activate the expression of RsmX and RsmY sRNA expression in chosen conditions. CrcZ and CrcY have both been downregulated (3.2 – 7.1 fold) 7 minutes after H<sub>2</sub>O<sub>2</sub> exposure and 60 minutes after NaCl addition. These sRNAs are upregulated in poor carbon sources or in stationary stage (Sonnleitner and Bläsi, 2014) to abrogate the catabolite repression, which was not needed in the chosen conditions.

The RgsA RNA is reported to be important for the survival in oxidative stress in *P. syringae* and *P. fluorescens* (González *et al.*, 2008; Park *et al.*, 2013). In our data RgsA is slightly upregulated (1.9-fold) in oxidative stress T2. It could be that RgsA sRNA protects the cells in oxidative stress but its expression is not highly elevated when exposed to H<sub>2</sub>O<sub>2</sub> as it has been reported that its over-expression did not help in survival when exposed to H<sub>2</sub>O<sub>2</sub> in *P. fluorescens*. (González *et al.*, 2008). The other reason could also be that its maximal level of expression was missed because it occurs before 7 minutes or between 7 and 60 minutes. ErsA/spf/Spot42-like sRNA was shown to be important for a better survival of *P. syringae* and is induced under exposure to oxidative stress (Ferrara *et al.*, 2014; Park *et al.*, 2014). ErsA was 14.8-fold upregulated

in osmotic stress after 60 minutes but no change in expression is observed in oxidative stress in present dataset.

Two annotated sRNAs, P1 and P6, detected in a previous study on *P. putida* KT2440 were not detected here. In the earlier study 14 possible novel sRNAs were predicted and named according to the intergenic region (IGR) they were located in (Frank *et al.*, 2011). Of these only 5 were detected here (c4 antisense RNA 4, Pit104, Pit132, Pit140, and Pit148). For the rest there were no expression profiles in the present dataset. There are several possible explanations for why all the annotated sRNAs were not detected here including: (1) different cDNA library construction methods lead to different transcripts detected; (2) some RNAs may be defiant to reverse transcription in the cDNA library construction and be underrepresented in the final dataset (Gómez-Lozano *et al.*, 2012); (3) the detection method (Rockhopper) did not detect some transcripts (as it did not find 6S RNA, which had extremely high expression); (4) certain sRNAs are expressed only in certain conditions. One example is the characterized sRNA NrsZ in *P. aeruginosa*, which was not expressed here, although sequence homology to NrsZ is present in the *P. putida* KT2440 genome (Wenner *et al.*, 2014). It could be that this sRNA is not expressed in the studied conditions or is underrepresented in cDNA libraries prepared in this study. Moreover, *P. putida* KT2440 does not have *rhlA* gene, which is its target mRNA of NrsZ in *P. aeruginosa*, so if NrsZ is expressed in *P. putida* KT2440 and bears a biological function, it is not identical to that in *P. aeruginosa*.

In *P. aeruginosa* 232 and 380 *cis*-encoded RNAs have been detected in different studies (Wurtzel *et al.*, 2012; Gómez-Lozano *et al.*, 2014), and in *P. syringae* 124 genes had antisense transcription (Filiatrault *et al.*, 2010). This is the first report of *cis*-encoded RNA in *P. putida* KT2440, with 217 asRNAs detected. The numbers of genes having antisense

transcripts or antisense transcription in other organisms are ranging from 2-46% (Thomason and Storz, 2010 and references there within). In a recent study in *P. putida* KT2440 transcription starts (TSS) 36% of genes had antisense TSS (D'Arrigo *et al.*, 2016), while in this our study antisense transcripts were only found to 3.3% of the genes. The reason for the discrepancy could be that the study included all TSS starting within 100 bp in front or after the start or stop codon, respectively and in this study only those that are *cis*-encoded to the ORF of the gene were included. Additional possible reasons are described above.

Nine clusters of sRNAs reported in this study have been identified depending on their expression in different conditions. Significant changes of expression hint at these transcripts performing a possible regulatory role that is beneficial to the cells in adaptation to stressful conditions (Rau *et al.*, 2015). A total of 196 sRNAs showed differential expression in one stress condition, and 64 were differentially changed in at least 2 of the chosen conditions, suggesting they could be involved in a more general stress response.

It is yet unclear how many of these newly found sRNA candidates carry physiological functions *in vivo*. They could function as protein (Duss *et al.*, 2014) or even RNA sponges (Lalaouna *et al.*, 2015; Miyakoshi *et al.*, 2015). Sometimes small RNA transcripts can encode small proteins (Silby and Levy, 2008) or represent noise transcription in bacteria (Georg and Hess, 2011; Thomason *et al.*, 2015) Although many small RNAs have been found in various microorganisms, only a subset has been functionally characterized. The numbers of novel sRNA transcripts will probably increase in the following years with new conditions and cDNA library preparations tested.

### **Concluding remarks**

This study is to our knowledge the only transcriptome study of *P. putida* KT2440 exposed to osmotic, oxidative, and imipenem stresses at two time points. Extensive genome-wide expression changes were observed when *P. putida* KT2440 was subjected to sodium chloride, hydrogen peroxide and antibiotic imipenem. These conditions resulted in a dramatic increase and decrease of mRNA levels of genes functioning at preventing, counteracting or repairing the damage. Rapid inductions and repressions of abundance of genes achieved within only a few minutes after the exposure to the stress conditions demonstrate the fast and complex adaptation of *P. putida* KT2440. At the same time the difference in 7 to 60 minutes in transcript levels after the induction of the stress shows time-regulated processes and differences between the immediate and longer exposure to stress. The data collected here confirm many stress-induced responses described previously and add new insights due to the depth of the RNA-seq data.

A total of 440 sRNA transcripts were detected, which drastically increases the numbers of sRNA in *P. putida*, and adds knowledge on antisense RNAs not described previously in this organism. Differential regulation of sRNAs in different stress conditions provides clues to their possible regulatory roles, which can aid in selecting relevant transcripts for functional characterization. Revealing their regulatory mechanisms will yield insights into bacterial stress response mechanisms developed to adapt changing environmental conditions. Depending on their role, their overexpression or deletion may have useful applications in biotechnology to improve stress tolerance.

## Experimental procedures

### *Bacterial strains, media and growth conditions*

The *P. putida* KT2440 strain (DSM6125) was cultivated in M9 medium (per liter:  $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ , 70 g;  $\text{KH}_2\text{PO}_4$ , 30 g;  $\text{NH}_4\text{Cl}$ , 10 g;  $\text{NaCl}$ , 5 g) supplemented with 0.5 % glucose and trace metals (per liter:  $\text{H}_3\text{BO}_3$ , 300 mg;  $\text{ZnCl}_2$ , 50 mg;  $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ , 30 mg;  $\text{CoCl}_2$ , 200 mg;  $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$ , 10 mg;  $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$ , 20 mg; and  $\text{NaMoO}_4 \cdot 2\text{H}_2\text{O}$ , 30 mg) (Abril *et al.*, 1989) at 30°C and 250 rpm in this study, unless otherwise indicated.

Single colonies were grown overnight in 5 mL M9 medium and the cultures were diluted to a starting  $\text{OD}_{600}$  of 0.05 in 50 mL M9 medium in 250 mL shake flasks. At mid-exponential growth phase ( $\text{OD}_{600} \sim 0.6$ ) different compounds were added at different concentrations, followed by monitoring of growth ( $\text{OD}_{600}$ ) and survival. For osmotic stress, the following  $\text{NaCl}$  (Sigma) concentrations were tested: 0, 2, 3, 3.5, 4, 4.5, and 5 % (g/V). For oxidative stress, the following  $\text{H}_2\text{O}_2$  (Sigma) concentrations were tested: 0, 0.5, 1, 2, 5, 10, 15, 20, 25, and 30 mM. For membrane stress, the beta-lactam antibiotic imipenem (Sigma) was used and final concentrations of 0, 0.05, 0.1, 0.2, 0.4, and 0.8  $\mu\text{g/mL}$  were tested. For monitoring survival, 1 mL of the culture before or 1, 3 or 24 hours after compound addition were harvested. Colony forming units (CFU) were counted on LB Cm plates incubated at 30° C.

For RNA-seq experiments, the following compound concentrations were used: 3 %  $\text{NaCl}$ , 0.05 mM  $\text{H}_2\text{O}_2$  and 0.1  $\mu\text{g/mL}$  of imipenem. The cultures grown in the same manner as described above were harvested 7 and 60 minutes after the addition of the stress compounds and the control was a sample harvested just prior to compound addition. All experiments were carried out in 3 biological replicates.

### *Total RNA isolation*

RNA extraction was performed as previously described (Gómez-Lozano *et al.*, 2012). Briefly, 20 mL of harvested culture was mixed with 0.2 volumes of STOP solution (95% [v/v] ethanol, 5% [v/v] phenol). Cells were centrifuged, snap frozen and stored at -80° C. Total RNA was extracted with Trizol (Invitrogen) and treated with DNase I (Fermentas) for DNA removal. Total RNA integrity and quality were validated by Agilent 2100 Bioanalyzer (Agilent Technologies).

### *Library preparation and RNA sequencing*

Transcriptome libraries (LIB>100) were constructed as previously described (Gómez-Lozano *et al.*, 2012) with some modifications. The total RNA sample was depleted of rRNA with the Ribo Zero Kit for Gram Negative Bacteria (Illumina). cDNA libraries were prepared with the TruSeq Stranded mRNA Sample Preparation Kit (Illumina) following the Low Sample LS Protocol. Libraries were validated with a DNA 1000 chip on the Agilent 2100 Bioanalyzer and concentration was measured using a Qubit 2.0 Fluorometer (Invitrogen, Life Technologies). The concentration of each library was normalized to 10 nM in TE buffer and cDNA libraries were pooled together for sequencing on the Illumina HiSeq 2000 platform at Beckman Coulter Genomics. The transcriptome libraries were single-end sequenced with 100 bp reads.

### *Data analyses*

The RNA-seq data was trimmed using Trimmomatic (Bolger *et al.*, 2014) and analyzed with the open source software Rockhopper with the default settings and choosing reverse complement reads and strand specific analysis (McClure *et al.*, 2013) (version 2.0.3), which includes an automated pipeline of read mapping, normalization, quantification of



transcript abundance and sRNA identification. The reads were mapped to the sequenced reference *P. putida* KT2440 genome (GenBank accession no. NC\_002947.3). Using SAMtools (Li *et al.*, 2009) the mapped files were merged and the identification of novel transcripts was performed by visual inspection with Integrative Genomics Viewer (IGV) (Thorvaldsdottir *et al.*, 2013), as Rockhopper detects many false positives. Differential gene and sRNA expression analysis were carried out with the webserver T-REx (de Jong *et al.*, 2015) using the RPKM values generated in the Rockhopper analysis, where all the tested conditions were compared to the control – a sample harvested at the point of the addition of the compound.

### **Acknowledgements**

The authors thank Se Hyeuk Kim for the help with the figures, Anne de Jong and Martin Holm Rau for help with RNA-seq data. The work was supported by the Novo Nordisk Foundation and and PhD grants from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7-People-2012-ITN, under grant agreement No. 317058, "BACTORY".

### **Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

### **References**

Abril, M. a., Michan, C., Timmis, K.N., and Ramos, J.L. (1989) Regulator and enzyme specificities of the TOL plasmid-encoded upper pathway for degradation of aromatic hydrocarbons and expansion of the substrate range of the pathway. *J. Bacteriol.* **171**: 6782–6790.

- Albesa, I., Becerra, M.C., Battán, P.C., and Páez, P.L. (2004) Oxidative stress involved in the antibacterial action of different antibiotics. *Biochem. Biophys. Res. Commun.* **317**: 605–609.
- Aspedon, A., Palmer, K., and Whiteley, M. (2006) Microarray Analysis of the Osmotic Stress Response in *Pseudomonas aeruginosa*. *J. Bacteriol.* **188**: 2721–2725.
- Bagge, N., Bagge, N., Schuster, M., Schuster, M., Hentzer, M., Hentzer, M., et al. (2004) *Pseudomonas aeruginosa* Biofilms Exposed to Imipenem Exhibit Changes in Global Gene Expression and B-Lactamase and Alginate Production. *Antimicrob. Agents Chemother.* **48**: 1175–1187.
- Battesti, A., Majdalani, N., and Gottesman, S. (2011) The RpoS-Mediated General Stress Response in *Escherichia coli* \*. *Annu. Rev. Microbiol.* **65**: 189–213.
- Bignucolo, A., Appanna, V.P., Thomas, S.C., Auger, C., Han, S., Omri, A., and Appanna, V.D. (2013) Hydrogen peroxide stress provokes a metabolic reprogramming in *Pseudomonas fluorescens*: enhanced production of pyruvate. *J. Biotechnol.* **167**: 309–15.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.
- Brown, A.D. (1976) Microbial water stress. *Bacteriol. Rev.* **40**: 803–846.
- Capra, E.J. and Laub, M.T. (2012) The Evolution of Two-Component Signal Transduction Systems. *Annu. Rev. Microbiol.* **66**: 325–347.
- Castro, A.G. De, Bredholt, H., Strøm, A.R., and Tunnacliffe, A. (2000) Anhydrobiotic Engineering of Gram-Negative Bacteria Anhydrobiotic Engineering of Gram-Negative Bacteria. **66**: 4142–4144.
- Caswell, C.C., Oglesby-Sherrouse, A.G., and Murphy, E.R. (2014) Sibling rivalry: related bacterial small RNAs and their redundant and non-redundant roles. *Front. Cell. Infect. Microbiol.* **4**: 1–13.
- Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C.M., and Vogel, J. (2012) An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J.* **31**: 4005–4019.

Chavarría, M., Nikel, P.I., Pérez-Pantoja, D., and de Lorenzo, V. (2013) The Entner-Doudoroff pathway empowers *Pseudomonas putida* KT2440 with a high tolerance to oxidative stress. *Environ. Microbiol.* **15**: 1772–85.

Citron, M. and Schuster, H. (1990) The c4 repressors of bacteriophages P1 and P7 are antisense RNAs. *Cell* **62**: 591–598.

D'Arrigo, I., Bojanovič, K., Yang, X., Rau, M.H., and Long, K.S. (2016) Genome-wide mapping of transcription start sites yields novel insights into the primary transcriptome of *Pseudomonas putida*. *Environ. Microbiol.* (accepted manuscript).

Djonović, S., Urbach, J.M., Drenkard, E., Bush, J., Feinbaum, R., Ausubel, J.L., et al. (2013) Trehalose Biosynthesis Promotes *Pseudomonas aeruginosa* Pathogenicity in Plants. *PLoS Pathog.* **9**: e1003217.

Duss, O., Michel, E., Yulikov, M., Schubert, M., Jeschke, G., and Allain, F.H.-T. (2014) Structural basis of the non-coding RNA RsmZ acting as a protein sponge. *Nature* **509**: 588–592.

Dwyer, D.J., Kohanski, M.A., and Collins, J.J. (2009) Role of reactive oxygen species in antibiotic action and resistance. *Curr. Opin. Microbiol.* **12**: 482–489.

Farr, S.B. and Kogoma, T. (1991) Oxidative stress responses in *Escherichia coli* and *Salmonella typhimurium*. *Microbiol Rev* **55**: 561–85.

Fernández, M., Conde, S., De La Torre, J., Molina-Santiago, C., Ramos, J.L., and Duque, E. (2012) Mechanisms of resistance to chloramphenicol in *Pseudomonas putida* KT2440. *Antimicrob. Agents Chemother.* **56**: 1001–1009.

Ferrara, S., Brugnoli, M., de Bonis, A., Righetti, F., Delvillani, F., Dehò, G., et al. (2012) Comparative profiling of *Pseudomonas aeruginosa* strains reveals differential expression of novel unique and conserved small RNAs. *PLoS One* **7**: e36553.

Ferrara, S., Carloni, S., Fulco, R., Falcone, M., Macchi, R., and Bertoni, G. (2014) Post-transcriptional regulation of the virulence-associated enzyme AlgC by the  $\sigma(22)$ -dependent small RNA ErsA of *Pseudomonas aeruginosa*. *Environ. Microbiol.*

Filiatrault, M.J., Stodghill, P. V., Bronstein, P. a., Moll, S., Lindeberg, M., Grills, G., et al. (2010) Transcriptome analysis of *Pseudomonas syringae*

identifies new genes, noncoding RNAs, and antisense activity. *J. Bacteriol.* **192**: 2359–2372.

Frank, S., Klockgether, J., Hagendorf, P., Geffers, R., Schöck, U., Pohl, T., et al. (2011) *Pseudomonas putida* KT2440 genome update by cDNA sequencing and microarray transcriptomics. *Environ. Microbiol.* **13**: 1309–26.

Fu, X., Fu, N., Guo, S., Yan, Z., Xu, Y., Hu, H., et al. (2009) Estimating accuracy of RNA-Seq and microarrays with proteomics. *BMC Genomics* **10**: 161.

Galinski, E. a. and Trüper, H.G. (1994) Microbial behaviour in salt-stressed ecosystems. *FEMS Microbiol. Rev.* **15**: 95–108.

Georg, J. and Hess, W.R. (2011) *cis*-Antisense RNA, Another Level of Gene Regulation in Bacteria. *Microbiol. Mol. Biol. Rev.* **75**: 286–300.

Gómez-Lozano, M., Marvig, R.L., Molin, S., and Long, K.S. (2012) Genome-wide identification of novel small RNAs in *Pseudomonas aeruginosa*. *Environ. Microbiol.* **14**: 2006–16.

Gómez-Lozano, M., Marvig, R.L., Molina-Santiago, C., Tribelli, P.M., Ramos, J.-L., and Molin, S. (2015) Diversity of small RNAs expressed in *Pseudomonas* species. *Environ. Microbiol. Rep.* **7**: 227–236.

Gómez-Lozano, M., Marvig, R.L., Tulstrup, M.V., and Molin, S. (2014) Expression of antisense small RNAs in response to stress in *Pseudomonas aeruginosa*. *BMC Genomics* **15**: 783.

González, N., Heeb, S., Valverde, C., Kay, E., Reimmann, C., Junier, T., and Haas, D. (2008) Genome-wide search reveals a novel GacA-regulated small RNA in *Pseudomonas* species. *BMC Genomics* **9**: 167.

Gottesman, S. and Storz, G. (2011) Bacterial Small RNA Regulators : Versatile Roles and Rapidly Evolving Variations. *Cold Spring Harb. Perspect. Biol.* **1**: 1–16.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S.R. (2003) Rfam: An RNA family database. *Nucleic Acids Res.* **31**: 439–441.

Hishinuma, S., Yuki, M., Fujimura, M., and Fukumori, F. (2006) OxyR regulated the expression of two major catalases, KatA and KatB, along with

peroxiredoxin, AhpC in *Pseudomonas putida*. *Environ. Microbiol.* **8**: 2115–2124.

de Jong, A., van der Meulen, S., Kuipers, O.P., and Kok, J. (2015) T-REx: Transcriptome analysis webserver for RNA-seq Expression data. *BMC Genomics* **16**: 663.

Kets, E.P., Galinski, E. a, Wit, M. De, Bont, J. a De, Heipieper, H.J., Kets, E.P.W., et al. (1996) Mannitol, a novel bacterial compatible solute in *Pseudomonas putida* S12. *J. Bacteriol.* **178**: 6665–6670.

Kim, J. and Park, W. (2014) Oxidative stress response in *Pseudomonas putida*. *Appl. Microbiol. Biotechnol.* **98**: 6933–46.

Kopf, M., Klähn, S., Scholz, I., Hess, W.R., and Voß, B. (2015) Variations in the non-coding transcriptome as a driver of inter-strain divergence and physiological adaptation in bacteria. *Sci. Rep.* **5**: 9560.

Lalaouna, D., Carrier, M.-C., Semsey, S., Brouard, J.-S., Wang, J., Wade, J.T., and Massé, E. (2015) A 3' External Transcribed Spacer in a tRNA Transcript Acts as a Sponge for Small RNAs to Prevent Transcriptional Noise. *Mol. Cell* **58**: 393–405.

Landini, P., Egli, T., Wolf, J., and Lacour, S. (2014) sigmaS, a major player in the response to environmental stresses in *Escherichia coli*: Role, regulation and mechanisms of promoter recognition. *Environ. Microbiol. Rep.* **6**: 1–13.

Lee, Y., Peña-Llopis, S., Kang, Y.S., Shin, H.D., Demple, B., Madsen, E.L., et al. (2006) Expression analysis of the fpr (ferredoxin-NADP<sup>+</sup> reductase) gene in *Pseudomonas putida* KT2440. *Biochem. Biophys. Res. Commun.* **339**: 1246–1254.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Livny, J., Brenic, A., Lory, S., and Waldor, M.K. (2006) Identification of 17 *Pseudomonas aeruginosa* sRNAs and prediction of sRNA-encoding genes in 10 diverse pathogens using the bioinformatic tool sRNAPredict2. *Nucleic Acids Res.* **34**: 3484–3493.

- Mascher, T. (2013) Signaling diversity and evolution of extracytoplasmic function (ECF) sigma factors. *Curr. Opin. Microbiol.* **16**: 148–155.
- Matilla, M.A., Espinosa-Urgel, M., Rodríguez-Herva, J.J., Ramos, J.L., and Ramos-González, M. (2007) Genomic analysis reveals the major driving forces of bacterial life in the rhizosphere. *Genome Biol.* **8**: R179.
- McClure, R., Balasubramanian, D., Sun, Y., Bobrovskyy, M., Sumby, P., Genco, C. a, et al. (2013) Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res.* **41**: e140.
- Misra, H.S., Khairnar, N.P., Barik, A., Indira Priyadarsini, K., Mohan, H., and Apte, S.K. (2004) Pyrroloquinoline-quinone: A reactive oxygen species scavenger in bacteria. *FEBS Lett.* **578**: 26–30.
- Miyakoshi, M., Chao, Y., and Vogel, J. (2015) Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. *EMBO J.* **34**: 1478–1492.
- Molina-Santiago, C., Daddaoua, A., Fillet, S., Duque, E., and Ramos, J.L. (2014) Interspecies signalling: *Pseudomonas putida* efflux pump TtgGHI is activated by indole to increase antibiotic resistance. *Environ. Microbiol.* **16**: 1267–1281.
- Molina-Santiago, C., Daddaoua, A., Gómez-Lozano, M., Udaondo, Z., Molin, S., and Ramos, J.-L. (2015) Differential transcriptional response to antibiotics by *Pseudomonas putida* DOT-T1E. *Environ. Microbiol.* **17**: 3251–3262.
- Mukhopadhyay, A., He, Z., Alm, E.J., Arkin, A.P., Baidoo, E.E., Borglin, S.C., et al. (2006) Salt stress in *Desulfovibrio vulgaris* hildenborough: An integrated genomics approach. *J. Bacteriol.* **188**: 4068–4078.
- Nelson, K.E., Weinell, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Martins dos Santos, V. a P., et al. (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**: 799–808.
- Nielsen, D.R., Leonard, E., Yoon, S.-H., Tseng, H.-C., Yuan, C., and Prather, K.L.J. (2009) Engineering alternative butanol production platforms in heterologous bacteria. *Metab. Eng.* **11**: 262–73.

Nikel, P.I., Martínez-García, E., and de Lorenzo, V. (2014) Biotechnological domestication of pseudomonads using synthetic biology. *Nat. Rev. Microbiol.* **12**: 368–79.

Palma, M., DeLuca, D., Worgall, S., and Quadri, L.E.N. (2004) Transcriptome analysis of the response of *Pseudomonas aeruginosa* to hydrogen peroxide. *J. Bacteriol.* **186**: 248–52.

Park, S.H., Butcher, B.G., Anderson, Z., Pellegrini, N., Bao, Z., D'Amico, K., and Filiatrault, M.J. (2013) Analysis of the small RNA P16/RgsA in the plant pathogen *Pseudomonas syringae* pv. *tomato* strain DC3000. *Microbiology* **159**: 296–306.

Park, S.H., Zhongmeng, B., Butcher, B.G., D'Amico, K., Xu, Y., Stodghill, P., et al. (2014) Analysis of the small RNA spf in the plant pathogen *Pseudomonas syringae* pv. *tomato* strain DC3000. *Microbiology* **160**: 941–953.

Park, W., Peña-Llopis, S., Lee, Y., and Demple, B. (2006) Regulation of superoxide stress in *Pseudomonas putida* KT2440 is different from the SoxR paradigm in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **341**: 51–56.

Poblete-Castro, I., Becker, J., Dohnt, K., dos Santos, V.M., and Wittmann, C. (2012) Industrial biotechnology of *Pseudomonas putida* and related species. *Appl. Microbiol. Biotechnol.* **93**: 2279–90.

Poole, L.B. and Ellis, H.R. (1996) Flavin-dependent alkyl hydroperoxide reductase from *Salmonella typhimurium*. Purification and enzymatic activities of overexpressed AhpF and AhpC proteins. *Biochemistry* **35**: 56–64.

Rachid, S., Ohlsen, K., Wallner, U., Hacker, J., Hecker, M., Ziebuhr, W., and Hacker, R.G. (2000) Alternative Transcription Factor  $\zeta$  B Is Involved in Regulation of Biofilm Expression in a *Staphylococcus aureus* Mucosal Isolate  
Alternative Transcription Factor B Is Involved in Regulation of Biofilm Expression in a *Staphylococcus aureus* Mucosal Isolate. *Society* **182**: 1–4.

Ramos, J.-L., Sol Cuenca, M., Molina-Santiago, C., Segura, a., Duque, E., Gomez-Garcia, M.R., et al. (2015) Mechanisms of solvent resistance mediated by interplay of cellular factors in *Pseudomonas putida*. *FEMS Microbiol. Rev.* 555–566.

Ramos, J.L., Martí, M., Molina-henares, A.J., Tera, W., Brennan, R., and Tobes, R. (2005) The TetR Family of Transcriptional Repressors. *Microbiol Mol Biol Rev* **69**: 326–356.

Rau, M.H., Bojanovič, K., Nielsen, A.T., and Long, K.S. (2015) Differential expression of small RNAs under chemical stress and fed-batch fermentation in *E. coli*. *BMC Genomics* **16**: 1051.

Romantsov, T., Guan, Z., and Wood, J.M. (2009) Cardiolipin and the osmotic stress responses of bacteria. *Biochim. Biophys. Acta - Biomembr.* **1788**: 2092–2100.

Rüberg, S., Tian, Z.X., Krol, E., Linke, B., Meyer, F., Wang, Y., et al. (2003) Construction and validation of a *Sinorhizobium meliloti* whole genome DNA microarray: Genome-wide profiling of osmoadaptive gene expression. *J. Biotechnol.* **106**: 255–268.

Silby, M.W. and Levy, S.B. (2008) Overlapping protein-encoding genes in *Pseudomonas fluorescens* Pf0-1. *PLoS Genet.* **4**: 1–8.

Sonnleitner, E., Abdou, L., and Haas, D. (2009) Small RNA as global regulator of carbon catabolite repression in *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U. S. A.* **106**: 21866–21871.

Sonnleitner, E. and Bläsi, U. (2014) Regulation of Hfq by the RNA CrcZ in *Pseudomonas aeruginosa* carbon catabolite repression. *PLoS Genet.* **10**: e1004440.

Sonnleitner, E., Gonzalez, N., Sorger-Domenigg, T., Heeb, S., Richter, A.S., Backofen, R., et al. (2011) The small RNA PhrS stimulates synthesis of the *Pseudomonas aeruginosa* quinolone signal. *Mol. Microbiol.* **80**: 868–885.

Sonnleitner, E. and Haas, D. (2011) Small RNAs as regulators of primary and secondary metabolism in *Pseudomonas* species. *Appl. Microbiol. Biotechnol.* **91**: 63–79.

Storz, G., Vogel, J., and Wassarman, K.M. (2011) Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Mol. Cell* **43**: 880–891.

Thomason, M.K., Bischler, T., Eisenbart, S.K., Förstner, K.U., Zhang, A., Herbig, A., et al. (2015) Global transcriptional start site mapping using



differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. *J. Bacteriol.* **197**: 18–28.

Thomason, M.K. and Storz, G. (2010) Bacterial Antisense RNAs: How Many Are There, and What Are They Doing? *Annu. Rev. Genet.* **44**: 167–188.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**: 178–192.

Tsuzuki, M., Moskvina, O. V., Kuribayashi, M., Sato, K., Retamal, S., Abo, M., et al. (2011) Salt stress-induced changes in the transcriptome, compatible solutes, and membrane lipids in the facultatively phototrophic bacterium *Rhodobacter sphaeroides*. *Appl. Environ. Microbiol.* **77**: 7551–9.

Waters, L.S. and Storz, G. (2009) Regulatory RNAs in Bacteria. *Cell* **136**: 615–628.

Weber, A. and Jung, K. (2002) Profiling Early Osmostress-Dependent Gene Expression in *Escherichia coli* Using DNA Macroarrays. *Society* **184**: 5502–5507.

Weinberg, Z., Wang, J.X., Bogue, J., Yang, J., Corbino, K., Moy, R.H., and Breaker, R.R. (2010) Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol.* **11**: R31.

Wenner, N., Maes, A., Cotado-Sampayo, M., and Lapouge, K. (2014) NrsZ: a novel, processed, nitrogen-dependent, small non-coding RNA that regulates *Pseudomonas aeruginosa* PAO1 virulence. *Environ. Microbiol.* **16**: 1053–1068.

Wurtzel, O., Yoder-Himes, D.R., Han, K., Dandekar, A. a, Edelheit, S., Greenberg, E.P., et al. (2012) The single-nucleotide resolution transcriptome of *Pseudomonas aeruginosa* grown in body temperature. *PLoS Pathog.* **8**: e1002945.

Yeom, J., Imlay, J.A., and Park, W. (2010) Iron homeostasis affects antibiotic-mediated cell death in *Pseudomonas* species. *J. Biol. Chem.* **285**: 22689–22695.

Yeom, J., Lee, Y., and Park, W. (2012) ATP-dependent RecG helicase is required for the transcriptional regulator OxyR function in *Pseudomonas* species. *J. Biol. Chem.* **287**: 24492–24504.

Zhao, S., Fung-Leung, W.P., Bittner, A., Ngo, K., and Liu, X. (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One* **9**: e78644.

Zheng, M., Wang, X., Templeton, L.J., Dana, R., Larossa, R. a, Storz, G., et al. (2001) DNA Microarray-Mediated Transcriptional Profiling of the *Escherichia coli* Response to Hydrogen Peroxide DNA Microarray-Mediated Transcriptional Profiling of the *Escherichia coli* Response to Hydrogen Peroxide. *J. Bacteriol.* **183**: 4562.

## Supplementary Information

**Table S1.** Overview of mapping statistics for the RNA sequencing libraries used in this study.

**Table S2.** *Pseudomonas putida* KT2440 annotated sRNAs and candidate sRNAs with homologies in the Rfam database.

**Table S3.** Novel intergenic sRNA transcripts in *P. putida* KT2440 (Pit). (Not reported in this thesis as too long).

**Table S4.** Novel antisense sRNAs transcripts in *P. putida* KT2440 (Pat). (Not reported in this thesis as too long).

**Table S5.** Complementary sRNA transcripts in *P. putida* KT2440.

**Table S6.** Novel sRNA transcripts conserved in organisms outside the *Pseudomonadaceae* family. (Not reported in this thesis as too long).

**Table S7.** Homologous sRNAs transcripts in *P. putida* KT2440.

**Table S8.** Differentially expressed genes (fold change  $\geq 2$ , p-value  $\leq 0.05$ ) under osmotic stress conditions (T1 and T2). (Not reported in this thesis as too long).

**Table S9.** Differentially expressed genes (fold change  $\geq 2$ , p-value  $\leq 0.05$ ) under oxidative stress conditions (T1 and T2). (Not reported in this thesis as too long).

**Table S10.** Differentially expressed genes (fold change  $\geq 2$ , p-value  $\leq 0.05$ ) under imipenem stress conditions (T1 and T2). (Not reported in this thesis as too long).

**Table S11.** Differentially expressed sRNAs (fold change  $\geq 2$ , p-value  $\leq 0.05$ ) in multiple stress conditions. (Not reported in this thesis as too long).

**Table S1.** Overview of mapping statistics for the RNA sequencing libraries used in this study.

Condition	Number of biological replicates	Library name	Total number of reads	Number of mapped reads	Number of reads mapping to protein coding genes (%)	Number of reads mapping antisense to protein regions (%)	Number of reads mapping to rRNA (%)	Number of reads mapping to intergenic regions (%)
Exponential growth (control)	3	C_14_1	16,318,328	6696361 (41%)	17	1	25	14
		C_15_1	12,758,072	12587486 (99%)	28	1	32	35
		C_16_1	14,776,447	14671936 (99%)	39	1	27	29
H2O2 7 min	3	H2O2_9_1	10,230,319	9402826 (92%)	22	2	38	34
		H2O2_11_1	7,344,725	5472477 (75%)	24	2	35	37
		H2O2_12_1	8,416,513	5602970 (67%)	18	1	42	27
H2O2 60 min	3	H2O2_9_2	9,232,187	9179552 (99%)	30	2	27	37
		H2O2_11_2	21,055,607	17251667 (82%)	21	1	31	35
		H2O2_12_2	10,250,336	9144734 (89%)	21	2	33	38
Imipenem 7 min	3	IP_5_1	9,554,118	9493735 (99%)	30	1	32	30
		IP_7_1	11,684,566	11641443 (100%)	33	1	24	39
		IP_8_1	8,926,964	8896613 (10%)	31	1	30	35
Imipenem 60 min	3	IP_5_2	2,557,260	2518848 (98%)	28	1	26	43
		IP_7_2	6,985,922	6661039 (95%)	25	1	38	33
		IP_8_2	4,002,625	3960712 (99%)	20	1	51	25
NaCl 7 min	3	NaCl_1_1	11,575,480	10522745 (91%)	23	1	36	36
		NaCl_2_1	15,715,295	14079601 (90%)	26	1	32	35
		NaCl_3_1	11,906,584	10619489 (89%)	26	1	34	33
NaCl 60 min	3	NaCl_1_2	10,750,141	10668542 (99%)	39	3	30	24
		NaCl_2_2	11,697,539	11306426 (97%)	40	3	28	24
		NaCl_3_2	9,269,100	9190989 (99%)	47	3	23	24
Summary Average			225,008,128	199,570,191			32	

**Table S2.** *Pseudomonas putida* KT2440 annotated sRNAs and candidate sRNAs with homologies in the Rfam database.

Nr.	Name	Start	Stop	Length	Strand	Upstream flanking gene	Downstream flanking gene	Orientation	Rfam	Blas	Ref.	Cluster
1	Spot42-like/spf/ErsA	130362	130561	200	+	PP_0123	PP_0124	>><	Pseudomon-1	+	1, 2, 3, 4	4
2	gabT	264769	264873	105	+	PP_0213	PP_0214	>>>	gabT	+	4, 14	
3	c4 antisense RNA 1	335696	335870	175	+	PP_0277	PP_0278	<>>	C4	+	14	1
4	RsmY	450752	450916	165	+	PP_0370	PP_0371	>><	RsmY	+	2, 3, 4, 7, 8, 13	2
5	P26	537405	537502	98	+	PP_0446	PP_0447	>>>	P26	+	2, 3, 4, 8	
6	rpsL leader	546001	546170	170	+	PP_0448	PP_0449	>>>	rpsL_pseudo	+	10	
7	Alpha RBS	561399	561492	94	+	PP_0475	PP_0476	>>>			11	
8	FMN riboswitch	616507	616373	135	-	PP_0530	PP_0531	<<<	FMN	++	2, 3	
9	c4 antisense RNA 2	759513	759682	170	+	PP_0651	PP_0652	>><	C4	+	14	
10	YybP-YkoY	876097	875944	154	-	PP_0760	PP_0761	<<<		+	2, 3	2
11	PhrS	1316293	1316402	110	+	PP_1148	PP_1150	>>>	PhrS	+	5, 8, 13	1
12	2 group II 1	1425775	1425975	201	+	PP_1249	PP_1250	>>>	group-II-D1D4-3	++	6	2
13	RnpB/P28/RNase P RNA	1512683	1513072	390	+	PP_1326	PP_1328	>>>	RNaseP_bact_a	+	2, 3, 4, 5, 13	
14	Pseudomon-groES RNA	1549132	1549255	124	+	PP_1359	PP_1360	>>>	Pseudomon-groES	+	14	2
15	t44	1785119	1785225	107	+	PP_1590	PP_1591	<>>	t44	+	2, 3, 4	
16	RsmZ	1822011	1822181	171	+	PP_1624	PP_1625	>><	PrrB_RsmZ	+	2, 2, 4, 13	
17	Cobalamin riboswitch 1	1866975	1867159	185	+	PP_1671	PP_1672	<>>	Cobalamin	+	2, 3	2
18	gyrA	1970946	1970997	52	+	PP_1766	PP_1767	>>>		+	14	1
19	2 group II 2	2069323	2069493	171	+	PP_1845	PP_1846	>>>	group-II-	++	6	2

20	RgsA/P16	2229834	2229726	109	-	PP_1967	PP_1968	><>	D1D4-3 P16	+	2, 3, 5, 8	
21	c4 antisense RNA 3	2303002	2302769	234	-	PP_2026	PP_2027	<<<	C4	+	14	
22	rmf RNA motif	2388741	2388343	399	-	PP_2095	PP_2096	><>	rmf	+	3	
23	Cobalamin riboswitch 2	2765195	2765043	153	-	PP_2418	PP_2419	<<<	Cobalamin	+	14	2
24	c4 antisense RNA 6	2855911	2855757	155	-	PP_2507	PP_2508	><>	C4	+	14	
25	P15	3466252	3466082	171	-	PP_3080	PP_3081	<<<		+	3, 4, 8	
26	TPP riboswitch 1	3613951	3614033	83	+	PP_3184	PP_3185	<>>	TPP	++	6	2
27	Cobalamin riboswitch 3	3981922	3981816	107	-	PP_3508	PP_3509	<<>	Cobalamin	+	14	2
28	CrcY	4013165	4013581	417	+	PP_3540	PP_3541	>><	CrcZ	+	2, 3, 4	
29	PrrF2	4595123	4595325	203	+	PP_4069	PP_4070	>>>	PrrF	+	2, 4	
30	sucA-II RNA	4735743	4735637	107	-	PP_4189	PP_4190	<<<	sucA-II	+	14	
31	c4 antisense RNA 7	4856709	4856553	157	-	PP_4270	PP_4271	><>	C4	+	14	
32	Bacteria_small_ SRP/4.5S rRNA	4858513	4858392	122	-	PP_4273	PP_4274	><<	Bacteria_sma II_SRP	+	2	
33	c4 antisense RNA 4/IGR 4535	5149065	5148926	140	-	PP_4534	PP_4535	<<<	C4	+	14	
34	PrrF1	5325394	5325493	100	+	PP_4685	PP_4686	>><	PrrF	+	2, 3, 5	
35	CrcZ	5338210	5338622	413	+	PP_4696	PP_4697	>>>	CrcZ	+	2, 3, 4, 12	2
36	P30	5338614	5338287	328	-	PP_4696	PP_4697	><>	CrcZ (-)	+	4, 8, 13	1
37	P31	5373151	5373213	63	+	PP_4724	PP_4725	<<<	P31	+	3, 8	
38	P32	5373351	5373255	97	-	PP_4724	PP_4725	<<<		+	3, 8, 13	4
39	SsrA tmRNA	5389943	5390415	473	+	PP_4738	PP_4739	>>>	tmRNA	+	2, 3, 4, 13, 15	
40	c4 antisense RNA 5	5390629	5390766	138	+	PP_4738	PP_4739	>>>	C4	+	14	
41	P24	5437810	5437675	136	-	PP_4775	PP_4776	<<>	P24	+	2, 4	2
42	TPP riboswitch 2	5596316	5596174	143	-	PP_4922	PP_4923	<<>	TPP	+	2, 3	2

43	SAH riboswitch	5667848	5667999	152	+	PP_4975	PP_4976	< >	SAH_riboswit ch	+	4, 14	2
44	6S/SsrS	5934663	5934842	180	+	PP_5202	PP_5203	> >	6S	+	2, 3, 4	
45	Pseudomon- Rho	5948619	5948465	155	-	PP_5214	PP_5215	< <	Pseudomon- Rho	+	14	

Rfam - candidate sRNAs new queried against Rfam database and matches of known sRNAs are indicated.

BLASTN - the sequence conservation of candidate sRNAs in other microbial organisms was investigated using BLASTN algorithm: (+) sequence conservation primarily in Pseudomonadaceae; (+++) sequence conserved in bacterial species outside the Pseudomonadaceae family.

Cluster - number of the cluster from differential expression of the sRNA for significantly differentially expressed sRNA (for more info see Fig. 5)

Ref.

1	Ferrara, S. <i>et al.</i> (2014). Post-transcriptional regulation of the virulence-associated enzyme AlgC by the $\sigma(22)$ -dependent small RNA ErsA of <i>Pseudomonas aeruginosa</i> . <i>Environmental Microbiology</i> .
2	Filiatrault, M. J. <i>et al.</i> (2010). Transcriptome analysis of <i>Pseudomonas syringae</i> identifies new genes, noncoding RNAs, and antisense activity. <i>Journal of Bacteriology</i> , <b>192</b> (9).
3	Frank, S. <i>et al.</i> (2011). <i>Pseudomonas putida</i> KT2440 genome update by cDNA sequencing and microarray transcriptomics. <i>Environmental Microbiology</i> , <b>13</b> (5).
4	Gómez-Lozano, M. <i>et al.</i> (2012). Genome-wide identification of novel small RNAs in <i>Pseudomonas aeruginosa</i> . <i>Environmental Microbiology</i> , <b>14</b> (8).
5	González, N. <i>et al.</i> (2008). Genome-wide search reveals a novel GacA-regulated small RNA in <i>Pseudomonas</i> species. <i>BMC Genomics</i> , <b>9</b> .
6	Griffiths-Jones, S. <i>et al.</i> (2003). Rfam: An RNA family database. <i>Nucleic Acids Research</i> , <b>31</b> (1).
7	Kay, E. <i>et al.</i> (2006). Two GacA-Dependent Small RNAs Modulate the Quorum-Sensing Response in <i>Pseudomonas aeruginosa</i> . <i>Journal of Bacteriology</i> , <b>188</b> (16).
8	Livny, J. <i>et al.</i> (2006). Identification of 17 <i>Pseudomonas aeruginosa</i> sRNAs and prediction of sRNA-encoding genes in 10 diverse pathogens using the bioinformatic tool sRNAPredict2. <i>Nucleic Acids Research</i> , <b>34</b> (12).
9	Moreno, R. <i>et al.</i> (2012). Two small RNAs, CrcY and CrcZ, act in concert to sequester the Crc global regulator in <i>Pseudomonas putida</i> , modulating

- 
- catabolite repression. *Molecular Microbiology*, **83**.
- 10 Naville, M. *et al.* (2010). Premature terminator analysis sheds light on a hidden world of bacterial transcriptional attenuation. *Genome Biology*, **11**(9).
- 11 Schlax, P. J. *et al.* (2001). Translational repression of the *Escherichia coli* alpha operon mRNA. Importance of an mRNA conformational switch and a ternary entrapment complex. *Journal of Biological Chemistry*, **276**(42).
- 12 Sonnleitner, E. *et al.* (2009). Small RNA as global regulator of carbon catabolite repression in *Pseudomonas aeruginosa*. *PNAS*, **106**.
- 13 Sonnleitner, E. *et al.* (2008). Detection of small RNAs in *Pseudomonas aeruginosa* by RNomics and structure-based bioinformatic tools. *Microbiology*, **154**(Pt 10).
- 14 Weinberg, Z. *et al.* (2010). Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biology*, **11**(3).
- 15 Williams, K. P. *et al.* (1996). Phylogenetic analysis of tmRNA secondary structure. *RNA*.
-



**Table S5.** Complementary sRNA transcripts in *P. putida* KT2440.

Nr.	Name	Strand	Name	Strand
1	Pit032	-	Pit031	+
2	Pit128	-	Psr2/CrcY	+
3	Pit129	-	Psr2/CrcY	+
4	Pit157	-	SsrA tmRNA	+
5	Pit158	-	SsrA tmRNA	+
6	Pit063	-	RsmZ	+
7	Pit164	-	6S/SsrS	+
8	Pit146*	-	Pit167*	-
9	Pit020	-	RsmY	+
10	Pit019	-	Pit018	+
11	Pit038	-	Pit037	+
12	P24	-	Pat203	+
13	Pit046	-	Pit045	+
14	Pit178	-	Pit177	+
15	Pit130*	-	Pat180*	-
16	Pit071	-	Pit070	+
17	Pit003	-	Pit002	+
18	P30	-	CrcZ	+
19	Pit144	-	Prrf2	+
20	Pit176	-	Pit175	+
21	rmf	-	Pit090	+
22	SRP/4.5S rRNA	-	Pit145	+

\* These transcripts are antisense to each other but are encoded in a different genomic context.

**Table S7.** Homologous sRNAs transcripts in *P. putida* KT2440.

Homologous sRNAs are reported in the same line.

Nr.		Homologous sRNA transcripts							
1	Psr2/CrcX	CrcZ							
2	Prrf1	PrrF2							
9	2 group II 1	2 group II 2							
7	C4 AS RNA 3	C4 AS RNA 1							
3	Pit017	Pit126							
4	Pit024	Pit064	Pit092	Pit127	Pit153	Pit163	Pit169		
5	Pit105	Pit137	Pit049	Pit056	Pit124	Pit132	Pit154	Pit162	Pit106
6	Pit048	Pit055	Pit107	Pit125	Pit133	Pit155	Pit161		
8	Pit052	Pit051							
10	Pat121	Pat122	Pat123	Pat172					
11	Pat056	Pat127	Pat128	Pat129	Pat135	Pat200			
12	Pat207	Pat208							
13	Pat019	Pat026	Pat041	Pat059	Pat088	Pat139	Pat197		
14	Pat021	Pat028	Pat029	Pat043	Pat061	Pat090	Pat141	Pat195	
15	Pat017	Pat024	Pat039	Pat057	Pat086	Pat136	Pat199		
16	Pat015	Pat022	Pat037	Pat054	Pat084	Pat133	Pat202		
17	Pat083	Pat100	Pat154	Pat160	Pat193				
18	Pat018	Pat025	Pat040	Pat058	Pat087	Pat137	Pat138	Pat198	
19	Pat020	Pat027	Pat042	Pat060	Pat089	Pat140	Pat196		
20	Pat016	Pat023	Pat038	Pat055	Pat085	Pat134	Pat201		
21	Pat093	Pat094	Pat095						



## **Research Article 3**

**Investigation of *Pseudomonas putida* KT2440 transcriptome in different carbon sources reveals novelty in bacterial uptake systems**

**I. D'Arrigo, J. G. R. Cardoso, M. Rennig, N. Sonnenschein, M. J. Herrgård, K. S. Long. (2016). Manuscript in preparation.**



**Investigation of *Pseudomonas putida* KT2440 transcriptome in different carbon sources reveals novelty in bacterial uptake systems**

Isotta D'Arrigo<sup>1</sup>, João G. R. Cardoso<sup>1</sup>, Maja Rennig<sup>1</sup>, Nikolaus Sonnenschein<sup>1</sup>, Markus J. Herrgård<sup>1</sup>, Katherine S. Long<sup>1,2</sup>

<sup>1</sup>The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Allé 6, DK-2970 Hørsholm, Denmark

<sup>2</sup>Address correspondence to: KSL [kalon@biosustain.dtu.dk](mailto:kalon@biosustain.dtu.dk) Phone: +45 45258024 Fax: +45 45258001

Keywords: *Pseudomonas putida*, KT2440, citrate, ferulic acid, serine, modelling, transporter

Running title: Novel *P. putida* transporters identified with RNA-seq

## Summary

*Pseudomonas putida* is characterized by a versatile metabolism and stress tolerance traits that make the bacterium able to cope with different environmental conditions. In this work, the ability of *P. putida* KT2440 to grow in the presence of four sole carbon sources (glucose, citrate, ferulic acid, serine) was investigated by a RNA sequencing (RNA-seq) approach and genome-scale metabolic modeling. Transcriptomic data identified uptake systems for the four carbon sources, and candidates were experimentally confirmed for their involvement in the assimilation of the substrates by mutant strain growth. The OpdH and BenF-like porins were involved in citrate and ferulic acid uptake, respectively. The citrate transporter (encoded by PP\_0147) and the TctABC system were important for supporting cell growth with citrate as sole carbon source; PcaT and VanK were associated with ferulic acid uptake; and the ABC transporter AapJPQM was involved in serine transport. A genome-scale metabolic model of *P. putida* KT2440 was constructed to integrate and analyze the transcriptomic data, identifying and confirming the active catabolic pathways for each carbon source. This study reveals novel information about transporters that are essential for understanding the bacterial adaptation to different environments.

## Introduction

*Pseudomonas putida* is an ubiquitous, rod-shaped Gram-negative bacterium that colonizes several environmental niches (Nelson *et al.*, 2002; Timmis, 2002; Martins dos Santos *et al.*, 2004). Its versatile metabolism allows the bacterium to thrive in many different environments and to synthesize a broad range of compounds of industrial relevance, features that make *P. putida* an interesting platform for microbial cell factory development (Poblete-Castro *et al.*, 2012; Tiso *et al.*, 2014; Loeschcke and Thies, 2015). One of the best characterized *P. putida* strains is KT2440 (Regenhardt *et al.*, 2002), which is a plasmid-free derivative of *P. putida* mt-2 (Nakazawa, 2002) and the first Gram-negative soil bacterium to be certified as a biosafety strain by the Recombinant DNA Advisory Committee (Federal Register, 1982). The genome of KT2440 was sequenced in 2002 (Nelson *et al.*, 2002) and was recently re-annotated (Belda *et al.*, 2016).

*P. putida* KT2440 colonizes several ecosystems (e.g. soil, water and rhizosphere of plants), which are often characterized by rapid physiological and chemical changes (Palleroni and Moore, 1986; Espinosa-Urgel *et al.*, 2002). Therefore, the bacteria is able to exploit a wide variety of nutrients, and tolerate and persist in stressful conditions (Martins dos Santos *et al.*, 2004). Indeed, the *P. putida* KT2440 genome contains numerous genes related to the utilization of different carbon sources, which allows the bacterium to adapt to diverse environmental conditions (Jiménez *et al.*, 2002; Nelson *et al.*, 2002; Martins dos Santos *et al.*, 2004; Belda *et al.*, 2016). The bacterium preferably uses organic acids and amino acids, abundant in its own habitat, as sole carbon sources (La Rosa, Nogales, *et al.*, 2015). However, sugars and aromatic compounds, derived from plant materials, are also good sources of



energy for *P. putida* (Rojo, 2010; La Rosa, Behrends, Williams, Bundy, & Rojo, 2015; La Rosa, Nogales, et al., 2015).

Transcriptomic analysis by RNA sequencing (RNA-seq) technology facilitates the understanding of bacterial adaptation to different environmental conditions by highlighting differentially expressed genes and RNA regulatory mechanisms. The transcriptome of *P. putida* KT2440 growing in different conditions has been investigated in several studies and many bacterial features have been revealed, including regulatory RNAs (Frank *et al.*, 2011) and metabolic pathways (Kim *et al.*, 2013; Nikel *et al.*, 2013, 2015). Moreover, genome-scale metabolic models (GEMs) are also important tools in systems biology. These models describe the known metabolic portfolio of a given organism and how it is linked to the genotype (McCloskey *et al.*, 2013). Over the past decade, scientists have built and improved GEMs of *P. putida* KT2440, and these models have been used to investigate metabolism on a wide-range of substrates (Nogales *et al.*, 2008; Puchalka *et al.*, 2008; Sohn *et al.*, 2010; Oberhardt *et al.*, 2011; Belda *et al.*, 2016). However, during the most recent effort to improve these models, a discrepancy between the *in silico* predictions and *in vivo* growth data was identified (Belda *et al.*, 2016). Only 21% of the substrates supporting growth on BIOLOG plates were predicted correctly in the *in silico* analysis. This discrepancy is mostly caused by the lack of identified transporter proteins for specific compounds. By adding the required transport reactions to the model without specifying the underlying genes, the authors could predict 75% of the growth-enabling substrates (Belda *et al.*, 2016). This shows how important it is to identify and confirm transport systems in order to generate realistic models.

In this work, the transcriptome of *P. putida* KT2440 growing on four sole carbon sources, glucose, citrate, ferulic acid, and serine, was

explored. In order to investigate the changes in metabolism RNA-seq and genome-scale metabolic modeling were combined. The involvement of the most relevant transporters identified from the RNA-seq data was experimentally verified using transposon mutants of genes encoding porins and inner membrane systems possibly involved in citrate, ferulic acid and serine transport. The compromised growth of the transposon mutants confirmed the involvement of the identified porins and transporters in the uptake of citrate, ferulic acid and serine. Using the transcriptomic data and experimental evidence the GEM of *P. putida* KT2440 was improved further.

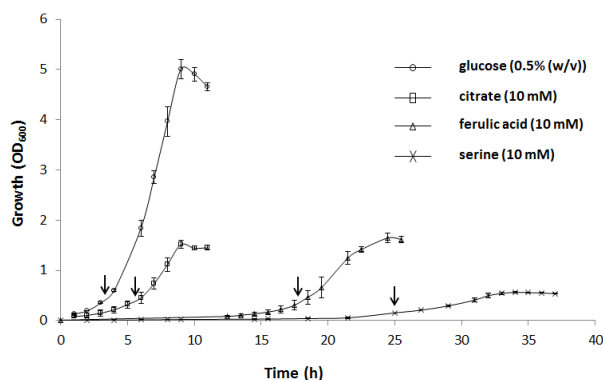
## Results

### *Growth of P. putida on four different sole carbon sources*

*P. putida* KT2440 was grown in minimal media supplemented with either glucose, citrate, ferulic acid or serine as sole carbon sources. The choice of the four compounds was mainly driven by the desire to explore different metabolic states resulting from the use of carbon sources entering at different points of central metabolism. Glucose is the preferred carbon source for many bacteria (e.g. *Escherichia coli* and *Bacillus subtilis*). However, it is not for *P. putida* (Rojo, 2010), which rather prefers organic acids (e.g. succinate) or amino acids (Vílchez *et al.*, 2000; Lugtenberg *et al.*, 2001; Revelles *et al.*, 2007; La Rosa, Behrends, *et al.*, 2015; La Rosa, Nogales, *et al.*, 2015). Still, *P. putida* can assimilate glucose (La Rosa, Nogales, *et al.*, 2015) and, when entering the carbon metabolism, it activates the glycolytic regime. Differently from glucose, citrate enters directly through the carbon metabolism in the tricarboxylic acid (TCA) cycle. With citrate as the carbon source the metabolism flows through the gluconeogenic route (Nogales *et al.*, 2008; Puchałka *et al.*, 2008). *P. putida* KT2440 can also

use a wide range of aromatic compounds as carbon sources. Ferulic acid is processed through the protocatechuate pathway, entering in the TCA cycle through succinate and acetyl-CoA (Jiménez *et al.*, 2002). Finally, *P. putida* is able to grow with amino acids as carbon sources, and in particular, unlike *E. coli* (Raskó and Alföldi, 1971; Newman and Walker, 1982), it can grow on serine as sole carbon source (Nogales *et al.*, 2008).

The growth of *P. putida* KT2440 cells in the presence of the four carbon sources was monitored by measuring OD<sub>600</sub> at different time points (Fig. 1).



**Figure 1.** Growth curves of *P. putida* KT2440. Cells were grown in M9 minimal medium in the presence of glucose (0.5% (w/v)), citrate (10 mM), ferulic acid (10 mM), or serine (10 mM) as sole carbon sources. Cell growth was carried out at 30° C with shaking at 250 rpm, and monitored by measuring optical density at 600 nm. Growth rate in glucose  $\mu_{\text{MAX}}$  ( $\text{h}^{-1}$ )  $0.55 \pm 0.01$ , growth rate in citrate  $\mu_{\text{MAX}}$  ( $\text{h}^{-1}$ )  $0.40 \pm 0.05$ , growth rate in ferulic acid  $\mu_{\text{MAX}}$  ( $\text{h}^{-1}$ )  $0.36 \pm 0.01$ , growth rate in serine  $\mu_{\text{MAX}}$  ( $\text{h}^{-1}$ )  $0.16 \pm 0.01$ . Arrows indicate the cell harvest points. Error bars represent the standard deviations of three biological replicates.

The growth of cells on glucose was the fastest, with a doubling time of 1.25 h during the exponential phase. Moreover, cells grown on glucose reached the highest cell density upon entry into stationary phase ( $OD_{600} \sim 5$ ). Cells grown on citrate had a doubling time of 1.8 h in exponential phase and a final cell density of  $OD_{600} \sim 1.5$ . The cells grown on ferulic acid showed a 14 h lag phase, followed by exponential growth with a doubling time similar to citrate (almost 2 h). Cells grown on serine had the longest lag phase of 22 h and the slowest growth in exponential phase with a doubling time of 4.3 h.

#### *The experimental approach using RNA-seq*

*P. putida* KT2440 cells were harvested in exponential phase during growth on each of the four carbon sources (Fig. 1). Total RNA was extracted and used for preparation of strand-specific cDNA libraries. As depletion of rRNA was not performed on the total RNA samples, the final libraries were sequenced twice on the Illumina HiSeq platform to increase the number of generated reads. The merged sequencing resulted in 21 to 71 million reads per sample, where on average 90% of the reads mapped to the reference chromosome. Around 1.3 million (3%) of the mapped reads were related to protein-coding genes, and 3.1 million (7%) mapped to intergenic regions (Table S1). Although the majority of reads mapped to rRNA, the double sequencing allowed a satisfactory number of reads to be obtained for differential gene expression analysis (Creecy and Conway, 2015).

#### *Differential gene expression reveals insights into the metabolic routes and uptake systems of the four carbon sources*

The reads were mapped onto the *P. putida* KT2440 genome using the Rockhopper software (McClure *et al.*, 2013; Tjaden, 2015). The Reads

Per Kilobase Million (RPKM) gene counts were used as input to the T-REx pipeline (de Jong *et al.*, 2015) for differential expression analysis. In this work growth on glucose was considered as the control condition, as cells were grown easily and faster compared with the other three carbon sources (Fig. S1). Therefore, gene expression levels in citrate, ferulic acid and serine were compared with the glucose condition, and genes were considered differentially expressed when they exhibited a p-value  $\leq 0.05$  and fold change greater than or equal to 2.

### **Glucose**

Cells growing on glucose showed an upregulation of genes related to the transport of the sugar and assimilation by the Entner-Doudoroff (ED) pathway (Fig. S1A – red box). The data revealed that 47 genes were consistently upregulated in glucose versus the other three carbon sources (Table S2). The gene encoding the outer membrane porin OprB-1 (PP\_1019) was significantly induced. This result is in agreement with previous studies proposing the involvement of the porin in the transport of glucose to the periplasmic space in *P. aeruginosa* and *P. putida* (Saravolac *et al.*, 1991; Wylie and Worobec, 1995; del Castillo *et al.*, 2007). The genes encoding an ABC transporter (PP\_1015-18), the gluconate transporter GntP (PP\_3417), and the 2-ketogluconate transporter KguT (PP\_3377) were also highly expressed, as glucose can be transported into the cytoplasm directly, or alternatively as gluconate or 2-ketogluconate after oxidation in the periplasmic space (Martins dos Santos *et al.*, 2004; del Castillo *et al.*, 2007; Nikel *et al.*, 2015). Related to that, the dehydrogenases genes (PP\_3382, PP\_3383, PP\_3384), involved in the periplasmic oxidation of glucose to 2-ketogluconate (del Castillo *et al.*, 2007, 2008), were also upregulated. The genes encoding the three cytoplasmic kinase enzymes GlK (PP\_1011), GnuK (PP\_3416), KguK (PP\_3378) catalyzing the phosphorylation glucose to glucose-6-

phosphate, gluconate to gluconate-6-phosphate, and 2-ketogluconate to 2-ketogluconate-6-phosphate, respectively, were induced in glucose growth. The *zwf-1* (PP\_1022) and *pgl* (PP\_1023) genes, responsible for the conversion of glucose-6-phosphate to gluconate-6-phosphate, and *kguD* (PP\_3376), that reduces 2-ketogluconate-6-phosphate to gluconate-6-phosphate, were upregulated in this condition. The key ED genes, *edd* (PP\_1010) and *eda* (PP\_1024) were also upregulated. These genes are required to convert gluconate-6-phosphate to glyceraldehyde-3-phosphatase and pyruvate. Finally the genes encoding the pyruvate carboxylase *AccC-2* (PP\_5347) and the oxaloacetate decarboxylase *OadA* (PP\_5346) were also induced in growth on glucose. The upregulation of *accC-2* confirms the presence of a strong pyruvate shunt to produce oxaloacetate rather than the direct oxidation of malate in the TCA cycle, and the induction of *oadA* is required to counteract the excess oxaloacetate during glucose growth (del Castillo *et al.*, 2007; Sudarsan *et al.*, 2014).

### ***Citrate***

A total of 391 genes were differentially expressed (265 upregulated and 126 downregulated; see Table S3) when comparing growth on citrate with growth on glucose. Out of these 206 (157 up- and 49 downregulated) exhibited high expression variations with fold changes > 5. A total of thirty-four genes were related to transporters, of which twenty-two were upregulated on citrate (Table 1).

**Table 1.** Transporter-related genes upregulated in citrate versus glucose. The 22 transporter-related genes upregulated in citrate versus glucose (p-value  $\leq 0.05$ ). GeneID, fold change, gene name and function are reported.

GeneID	Fold	Name	Function
PP_0146	3.2	-	integral membrane protein TerC
PP_0147*	189.6	-	citrate transporter
PP_0296	5.5	-	glycine/betaine ABC transporter substrate-binding protein
PP_0805	25.6	-	TolC family type I secretion outer membrane protein
PP_0806	63.9	-	surface adhesion protein
PP_0962	2.2	ttg2E	toluene-tolerance protein
PP_1002	31.1	arcD	arginine/ornithine antiporter
PP_1070	3.3	-	polar amino acid ABC transporter inner membrane subunit
PP_1400	20.1	-	metabolite/H <sup>+</sup> symporter major facilitator superfamily
PP_1416*	22.9	-	tricarboxylate transport protein TctA
PP_1417*	20.1	-	tricarboxylate transport protein TctB
PP_1418*	66.6	-	tricarboxylate transport protein TctC
PP_1419*	91.2	-	porin
PP_1726	24.6	-	ABC transporter substrate-binding protein
PP_2628	3.3	-	ABC transporter ATP-binding protein
PP_3636	22.9	-	sulfonate ABC transporter substrate-binding protein
PP_3637	20.1	-	sulfonate ABC transporter ATP-binding protein
PP_4035	22.9	-	NCS1 nucleoside transporter
PP_4282	9.7	-	aquaporin Z
PP_4735	83	lctP	L-lactate transport
PP_4921	5.3	-	NCS1 nucleoside transporter
PP_5174	20.1	-	RND efflux transporter

\* Genes reported in the text as involved in citrate uptake.

Among the highly expressed genes were a predicted porin (PP\_1419) and the tricarboxylate transporter TctABC (PP\_1418-16) suggesting

their putative role in citrate transport. The predicted porin shares a 65% amino acid similarity with the OpdH porin of *P. aeruginosa* (PA0755), which has been previously implicated in the uptake of *cis*-aconitate (Tamber *et al.*, 2006) and in the uptake of other tricarboxylates, such as isocitrate and citrate (Tamber *et al.*, 2007). Therefore, the PP\_1419 gene product will subsequently be referred to here as OpdH. Moreover, the *opdH* gene has the same genomic organization in *P. aeruginosa* and *P. putida*, with the porin gene located directly upstream of the tricarboxylate transporter cluster *tctABC*. The *tctABC* gene cluster encodes the TctABC system consisting of a periplasmic tricarboxylate-binding protein TctC, and two cytoplasmic membrane proteins TctA and TctB. The three subunits have not yet been characterized in either *Pseudomonas* species (58% to 89% amino acid similarity between the two species). However, the TctABC transporter has been functionally characterized in *Salmonella enterica* serovar Typhimurium, where the TctC subunit has been shown to bind fluorocitrate, citrate, isocitrate and *cis*-aconitate (Sweet *et al.*, 1979, 1984; Ashton *et al.*, 1980; Somers *et al.*, 1981; Widenhorn *et al.*, 1988; Winnen *et al.*, 2003). The TctABC system of *S. typhimurium* was the first characterized member of the Tripartite Tricarboxylate Transporter (TTT) family. In *Salmonella*, the induction of the transporter during growth on tricarboxylate, is controlled by the two-component regulatory system TctED, encoded by genes contiguous to the *tctABC* cluster (Widenhorn *et al.*, 1988, 1989). Similarly in *P. putida* the *tctED* genes (PP\_1420-21) can be found nearby the *opdH-tctABC* gene cluster.

The gene encoding the putative citrate transporter (PP\_0147) was also highly induced in cells growing on citrate. Previous studies suggest that this gene is relevant for citrate uptake in *P. putida* (del Castillo *et al.*, 2007; Wang and Nomura, 2010). The citrate transporter belongs to the



CitMHS family of transporters, whose prototypes CitM and CitH were first identified in *B. subtilis* and shown to bind citrate (Boorsma *et al.*, 1996; Krom *et al.*, 2000).

Genes involved in the TCA cycle (isocitrate dehydrogenase *icd*/PP\_4011, succinate dehydrogenase *sdhD*/PP\_4192, malate:quinone oxidoreductases *mgo-3*/PP\_2925, isocitrate lyase *aceA*/PP\_4116), beta-oxidation (*fadB*/PP\_2136 and *fadE*/PP\_1893), and lipid metabolism (PP\_1689, PP\_4021, PP\_4817, PP\_5003) were also upregulated during growth on citrate (Fig. S1D – green box).

### ***Ferulic acid***

During growth on ferulic acid, 237 genes (145 upregulated and 92 downregulated) were differentially expressed compared with glucose (Table S4). Of these, 141 genes (98 upregulated and 43 downregulated) showed expression changes of more than 5-fold. The genes encoding the feruloyl-CoA synthase Fcs (PP\_3356), the *p*-hydroxycinnamoyl CoA hydratase/lyase Ech (PP\_3358), and the vanillin dehydrogenase Vdh (PP\_3357), responsible for the conversion of ferulic acid to vanillate, and the gene for the VanAB enzymes (PP\_3736-37) to transform vanillate into protocatechuate, were highly induced (Fig. S1C – purple box). These findings agree with previous studies describing ferulic acid metabolism via a non- $\beta$ -oxidative pathway (Gasson *et al.*, 1998; Narbad and Gasson, 1998; Venturi *et al.*, 1998; Overhage *et al.*, 1999; Jiménez *et al.*, 2002; Plaggenborg *et al.*, 2003; Calisti *et al.*, 2008; Graf and Altenbuchner, 2014). The PP\_3354 and *aat*/PP\_3355 genes were also upregulated and have been suggested as an alternative route by Overhage *et al.* (Overhage *et al.*, 1999) (Fig. S1C – purple box).

Interestingly, twenty-two transporter-related genes were differentially expressed, of which fifteen were upregulated in cells growing on ferulic acid (Table 2).

**Table 2.** Transporter-related genes upregulated in ferulic acid acid versus glucose. The 15 transporter-related genes upregulated in citrate versus glucose (p-value  $\leq 0.05$ ). GeneID, fold change, gene name and function are reported.

GeneID	Fold	Name	Function
PP_0179	27.4	-	HlyD family type I secretion membrane fusion protein
PP_0806	27.4	-	surface adhesion protein
PP_0962	2.5	ttg2E	toluene-tolerance protein
PP_1376*	271.1	pcaK	benzoate transport
PP_1378*	127.3	pcaT	metabolite/H <sup>+</sup> symporter major facilitator superfamily
PP_1383*	183	-	BenF-like porin
PP_1400	48	-	metabolite/H <sup>+</sup> symporter major facilitator superfamily
PP_2553	36.2	-	major facilitator family transporter
PP_2604	19.1	-	major facilitator family transporter
PP_2667	27.4	-	ABC transporter
PP_2668	21.6	-	ABC transporter ATP-binding protein
PP_3349	209.5	-	3-hydroxyphenylpropionic transporter MhpT
PP_3740*	271.1	-	major facilitator family transporter
PP_4035	24.5	-	NCS1 nucleoside transporter
PP_4707	19.5	-	transport-associated protein

\* Genes reported in the text as involved in ferulic acid acid uptake.

The gene encoding the BenF-like porin (PP\_1383), the transporter encoded by PP\_3740, the PcaK transporter (PP\_1376), and the PcaT metabolite/H<sup>+</sup> symporter (PP\_1378), were significantly induced during growth on ferulic acid, and may be involved in the transport of the compound into the cells. The *benF*-like porin gene (PP\_1383) is one of the twenty *benF*-like genes related to the benzoate-specific porin gene

*benF* (PP\_3168) in *P. putida* KT2440. The annotated *benF* gene is part of an operon involved in benzoate catabolism and regulated by *benR*. It has been proposed that the BenF porin may contribute to the flux of benzoate through the membrane (Cowles *et al.*, 2000; Nishikawa *et al.*, 2008; Parthasarathy *et al.*, 2010).

The PP\_3740 gene (here referred as *vanK* gene because of its homology with *vanK* from *Acinetobacter* sp. ADP1 (Jiménez *et al.*, 2002)), as well as the *pcaK* and *pcaT* genes, encode predicted inner membrane transport proteins from the aromatic catabolic pathways of *P. putida* KT2440, and they belong to the major facilitator superfamily (MFS) of transporters (Jiménez *et al.*, 2002). The *vanK* gene (PP\_3740) is localized near the *vanAB* genes (PP\_3736-37), while the *pcaK* and *pcaT* genes are included in a gene cluster (from PP\_1376 to the *benF*-like porin PP\_1383) that was upregulated in the present data. The cluster also includes: the *pcaF* (PP\_1377) gene encoding the beta-ketoadipyl CoA thiolase, the *pcaT* (PP\_1378) symporter gene, three metabolic genes *pcaB* (PP\_1379), *pcaD* (PP\_1380), *pcaC* (PP\_1381), and PP\_1382 encoding a hypothetical protein. Another two gene clusters of *pca* genes, *pcaGH* (PP\_4655-56) and *pcaIJ* (PP\_3951-52), were also upregulated. These *pca* gene clusters are involved in the aromatic catabolic pathway of protocatechuate in *P. putida* KT2440 leading to succinyl-CoA and acetyl-CoA (Harwood *et al.*, 1994; Harwood and Parales, 1996; Nichols and Harwood, 1997; Parke *et al.*, 2000, 2001; Jiménez *et al.*, 2002).

### ***Serine***

Cells grown on serine showed 928 differential expressed genes (557 upregulated and 371 downregulated) (Table S5) when compared with cells grown on glucose. Out of these, 458 (342 upregulated and 116 downregulated) show a variation higher than 5-fold. Fifty-six

transporter-related genes were differentially expressed and forty-three of them were upregulated in cells growing on serine (Table 3).

**Table 3.** Transporter-related genes upregulated in serine versus glucose. The 43 transporter-related genes upregulated in serine versus glucose (p-value  $\leq 0.05$ ). GeneID, fold change, gene name and function are reported.

GeneID	Fold	Name	Function
PP_0026	4.5	-	CDF family cobalt/cadmium/zinc transporter
PP_0057	18.2	-	major facilitator family transporter
PP_0294	20.3	-	glycine betaine/L-proline ABC transporter ATP-binding subunit
PP_0296	6.8	-	glycine/betaine ABC transporter substrate-binding protein
PP_0412	4.2	-	polyamine ABC transporter substrate-binding protein
PP_0713	24.6	-	potassium efflux system protein
PP_0805	46.1	-	TolC family type I secretion outer membrane protein
PP_0806	33.2	-	surface adhesion protein
PP_0907	24.6	-	RND efflux membrane fusion protein-like protein
PP_1002	48.3	arcD	arginine/ornithine antiporter
PP_1003	31.1	-	arginine/ornithine antiporter
PP_1071	2.2	-	amino acid ABC transporter substrate-binding protein
PP_1137	2.6	braG	branched chain amino acid ABC transporter ATP-binding protein
PP_1141	2.8	braC	ABC transporter substrate-binding protein
PP_1206*	4.1	oprD	porin
PP_1297*	25	aapJ	amino acid ABC transporter substrate-binding protein
PP_1298*	22.5	aapQ	polar amino acid ABC transporter permease
PP_1300*	12	aapP	general amino acid ABC transporter ATP-binding protein
PP_1383	16	-	BenF-like porin
PP_1445	24.6	oprB-2	porin

PP_1486	5.6	-	polyamine ABC transporter substrate-binding protein
PP_1726	17.2	-	ABC transporter substrate-binding protein
PP_1741	6.6	-	glycine betaine ABC transporter substrate-binding protein
PP_2241	24.6	-	major facilitator family transporter
PP_2264	24.6	-	ABC transporter substrate-binding protein
PP_2569	26.8	-	major facilitator superfamily transporter
PP_2769	18.2	-	branched-chain amino acid ABC transporter permease
PP_3025	2	-	amino acid transporter LysE
PP_3176	28.9	-	major facilitator family transporter
PP_3558	13	-	glycine betaine ABC transporter substrate-binding protein
PP_3656	4.7	-	aromatic compound-specific porin
PP_4035	280.4	-	NCS1 nucleoside transporter
PP_4282	17.2	-	aquaporin Z
PP_4428	18.2	-	amino acid ABC transporter substrate-binding protein
PP_4707	8.1	-	transport-associated protein
PP_4863	16.2	-	branched chain amino acid ABC transporter ATP-binding protein
PP_4864	11.6	-	branched chain amino acid ABC transporter ATP-binding protein
PP_4865	28.9	-	inner-membrane translocator
PP_4866	31.1	-	inner-membrane translocator
PP_4867	13	-	ABC transporter substrate-binding protein
PP_5174	18.2	-	RND efflux transporter
PP_5175	20.3	-	RND family efflux transporter MFP subunit
PP_5341	18.2	-	polyamine ABC transporter substrate-binding protein

\* Genes reported in the text as involved in serine uptake.

The *oprD* gene (PP\_1206), as well as *aapJ* (PP\_1297), *aapQ* (PP\_1298), and *aapP* (PP\_1300) encoding the three components of the ABC transporter AapJQMP, were found to be upregulated on serine, suggesting their participation in amino acid transport. The porin OprD

shows a 51% amino acid similarity with the homologous porin in *P. aeruginosa* (PA0958) that has been shown to act as a specific channel for amino acids, small peptides and carbapenems (Trias and Nikaido, 1990). AapJQMP is a general polar amino acid transporter homologous to AapJQMP of *Rhizobium leguminosarum* (51% to 79% amino acid similarity), which can transport a wide range of amino acids (Walshaw and Poole, 1996). The *aapJQMP* operon consists of four genes encoding a periplasmic binding protein, permease, inner membrane protein and ATP binding subunit, respectively.

Interestingly, the gene encoding the porin OprB-2 (PP\_1445), homologous to OprB-1, was highly induced in serine-grown cells versus glucose-grown cells. Although the role of OprB-1 in carbohydrate transport has been demonstrated (Saravolac *et al.*, 1991; Wylie and Worobec, 1995; del Castillo *et al.*, 2007), no study has addressed the role of OprB-2. Moreover, the gene encoding the BenF-like porin (PP\_1383) appeared to also be induced in serine, although with a lower fold-change than in ferulic acid as mentioned previously. Additionally, the PP\_4867-4863 gene cluster was also overexpressed in serine. The locus consists of the putative periplasmic ABC transporter substrate-binding protein gene (PP\_4867) and the branched chain amino acid ABC transporter operon (PP\_4866-4863). The cluster products show high amino acid similarity (78% to 86%) to the homologous branched chain amino acid ABC transporter in *P. aeruginosa* (PA4913-4909), of which the two gene products PA4912 and PA4910 have been shown to be important for D-alanine and D-valine utilization (Johnson *et al.*, 2008).

Growth on serine as sole carbon source affected the expression of several metabolic genes related to serine utilization (Fig. S1B – blue box). The gene encoding the serine dehydratase Sda-3 (PP\_3144), for the conversion of serine in pyruvate, was highly induced in serine

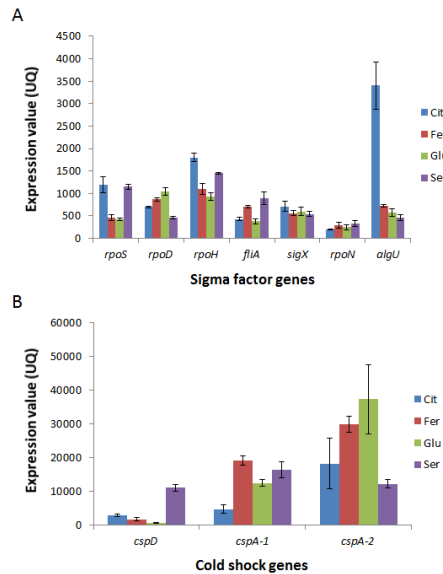
conditions, as well as serine O-acetyltransferase (PP\_3136). On the other hand other genes that use serine in biosynthetic pathways including the tryptophan synthase genes *trpA* (PP\_0082) and *trpB* (PP\_0083), and the serine/glycine hydroxymethyltransferase *glyA* (PP\_0671) gene were downregulated.

*Other differential expressed genes: sigma factors and cold-shock proteins*

The *rpoS* gene (PP\_1623), encoding the sigma factor related to stress and starvation in stationary phase (Ramos-González and Molin, 1998), was highly expressed in cells growing on citrate and serine (Fig. 2A). It has been shown that the level of the *rpoS* gene inversely correlates with the growth rate of *P. putida* KT2440 (Kim *et al.*, 2013). This can explain the low expression of *rpoS* on glucose and the high expression on serine compared with the other conditions. However, the *rpoS* level in citrate remained high compared with ferulic acid even though the two growth rates on the two carbon sources were very similar (Fig. 1). In connection with that, the expression of the gene encoding the cold shock protein CspD (PP\_4010) is inversely correlated with growth rates in *P. putida* (Kim *et al.*, 2013). Indeed, *cspD* had the highest expression level on serine and lowest on glucose (Fig. 2B). There are two other cold shock protein genes that were differentially expressed (Fig. 2B). The *cspA-1* gene (PP\_1522) was highly expressed during growth on ferulic acid, serine, and glucose, and weakly expressed in citrate. The *cspA-2* (PP\_2463) gene was generally upregulated in all four conditions.

A similar trend to the *rpoS* expression levels was observed for the *rpoH* gene, encoding the sigma factor that controls heat shock proteins. The *rpoH* gene (PP\_5108) was highly expressed in citrate and serine (Fig. 2A). Interestingly, the sigma factor AlgU (PP\_1427), which

negatively controls the expression of flagella in *P. aeruginosa* (Garrett *et al.*, 1999), and the response to osmotic stress and desiccation in *P. syringae* (Keith and Bender, 1999) and *P. fluorescens* (Schnider-Keel *et al.*, 2001) was upregulated 7-fold in citrate compared with the three other carbon sources (Fig. 2A). Finally, significant expression differences were not observed for the sigma factors genes *rpoD* (PP\_0387), *fliA* (PP\_4341), *sigX* (PP\_2088), and *rpoN* (PP\_0952) between the four conditions (Fig. 2A).



**Figure 2.** Expression of sigma factor and cold shock protein genes. The expression value of genes related to sigma factors (A), and genes encoding cold shock proteins (B), for each condition. The expression values obtained with Rockhopper analysis are normalized by the upper quartile of gene expression. The expression value is the average of thee replicates. Abbreviations: Glu, glucose; Cit, citrate; Fer, ferulic acid; Ser, serine.



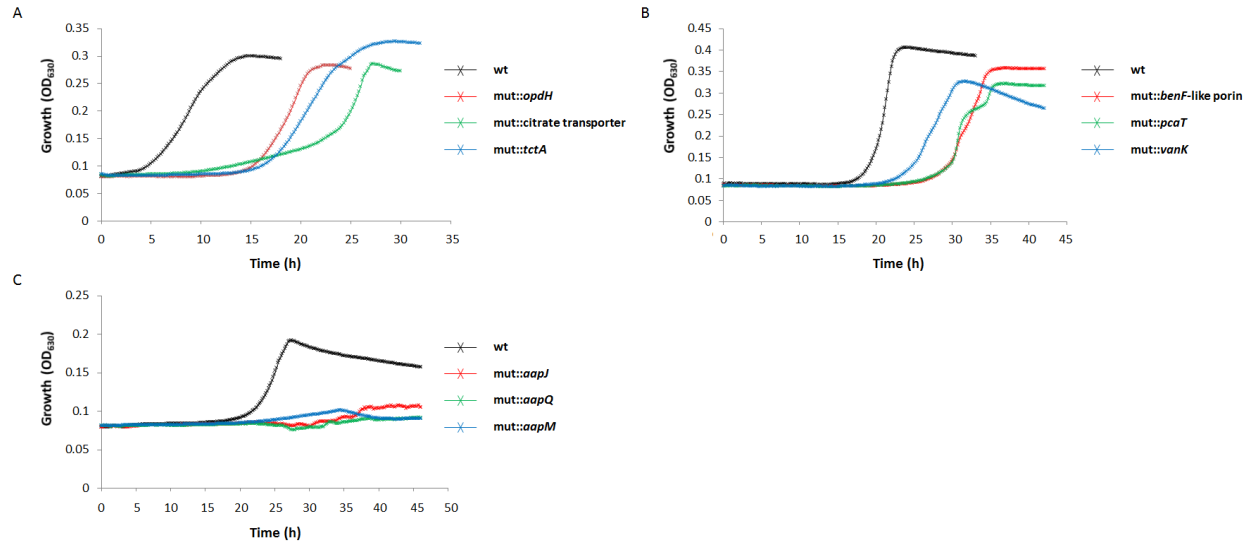
### *Metabolic modeling*

In the genome-scale metabolic model, three different systems can facilitate glucose uptake: a) transport of glucose via an ABC transporter and subsequent phosphorylation; b) periplasmic conversion of glucose into gluconate by NAD dehydrogenase followed by transport into the cytoplasm; and c) further periplasmic conversion of gluconate into 2-ketogluconate and then transport into the cytoplasm. The expression levels indicate that the glucose ABC transporter is highly expressed and the model simulations describe the same mechanism of transport (Fig. S2). The simulations show citrate being taken up through the citrate transporter (PP\_0147). The tricarboxylate transport system TctABC was also included in the model. The genes displayed high expression when cells were grown on citrate as sole carbon source. According to the simulations, citrate uptake can occur via either system (Fig. S3). Using the ChEBI database (de Matos *et al.*, 2009) we identified seven more tricarboxylic acids in the model; 2-methylcitrate, oxalosuccinate, methylisocitrate, 4-carboxy-2-hydroxyhexa-2,4-dienedioate, isocitrate, *cis*-2-methylaconitate and 3-carboxy-*cis,cis*-muconate. According to the model predictions, *P. putida* should be able to grow on all of them, with the exception of 4-carboxy-2-hydroxyhexa-2,4-dienedioate. We included the pathway for ferulic acid uptake and degradation in the model. The pathway contains two different branches for ferulic acid degradation as has been previously suggested (Overhage *et al.*, 1999). The constraints imposed in the model allow flux to be carried by either branch. This result is coherent with the expression data (Fig. S4). Finally, the expression data for serine highlight the transport of the amino acid via the ABC transporter (PP\_1297, PP\_1298, PP\_1299 and PP\_1300). Predicted fluxes and gene expression show that serine is directly converted into pyruvate (Fig. S5).

*Transposon mutants exhibit delayed growth on sole carbon sources*

To confirm the involvement of the identified porins and inner membrane uptake systems in carbon source assimilation, the mini-Tn5 transposon mutant strains of *P. putida* KT2440 were tested in a microtiter plate growth assay. Nine transposon mutants (defined as *mut::interrupted gene*) (Table S6) were obtained from the available collection at the Pseudomonas Reference Culture Collection (PRCC) (Duque *et al.*, 2007; Fernández *et al.*, 2012). The mutant strains examined have insertions in a gene of interest possibly involved in citrate, ferulic acid or serine uptake including *opdH* (PP\_1419), citrate transporter (PP\_0147), *tctA* (PP\_1416), *benF*-like porin (PP\_1383), *pcaT* (PP\_1378), *vanK* (PP\_3740), *aapJ* (PP\_1297), *aapQ* (PP\_1298), and *aapM* (PP\_1299).

All mutants show growth curves (rates and yields) similar to those of the KT2440 wild type (wt) strain when grown in presence of glucose as sole carbon source (data not shown), confirming the expression data and that the tested uptake systems are not involved in glucose assimilation. However, all mutants show compromised growth compared with KT2440 wt when grown in the presence of the specific carbon sources (citrate, ferulic acid or serine) associated with the disrupted gene (Fig. 3).



**Figure 3.** Transposon mutants growth assay. The transposon mutant strains (*mut::interrupted gene*) are grown in the specific carbon source of the disrupted gene possibly involved with. Growths are monitored by measuring optical density at 630 nm each 10 min in a kinetic assay in microtiter plate. Growth curves are the result of the average of three biological replicates. Growth on citrate (A), in ferulic acid (B), and in serine (C).

The mutants with insertions in the genes *opdH* (PP\_1419), citrate transporter (PP\_0147), and *tctA* (PP\_1416) showed defects (long lag phase) compared with KT2440 wt when grown with citrate as sole carbon source. The longer lag phase was observed for the mutant with a disrupted citrate transporter (PP\_0147) gene (Fig. 3A). The three mutants with insertions in genes associated with ferulic acid uptake, *benF*-like porin (PP\_1383), *pcaT* (PP\_1378), and *vanK* (PP\_3740), also showed a long lag phase compared with KT2440 wt during growth with ferulic acid as sole carbon source. Larger effects were observed for strains with disrupted *benF*-like porin and *pcaT* genes compared to the strain with a disrupted *vanK* gene (Fig. 3B). Transposon mutants with an insertion in one of the subunit genes for the ABC transporter AapJQMP exhibited compromised growth on serine (Fig. 3C).

The observed long lag phase compared with KT2440 wt showed that the absence of each specific porin or inner membrane transporter affects growth. This suggests that all the uptake systems identified here are involved in supporting growth on the specific carbon source. Moreover, additional porins in the outer membrane may be involved in the transport.

## Discussion

To gain insights into the metabolic versatility of *P. putida*, RNA-seq experiments were performed on cells growing on glucose, citrate, ferulic acid and serine as sole carbon sources. The four substrates are particularly interesting, as they require distinct utilization pathways delineating some features of the bacterial adaptation machinery to different environments. The results presented here, have clearly defined the central pathways for the assimilation of the four carbon sources.

*P. putida* KT2440 lacks the 6-phosphofructokinase enzyme and owns an incomplete glycolytic pathway for hexose assimilation. However it is able to assimilate glucose by the ED route (Nelson *et al.*, 2002; Martins dos Santos *et al.*, 2004; Fuhrer *et al.*, 2005; del Castillo *et al.*, 2007; Velázquez *et al.*, 2007; Chavarría *et al.*, 2013). In contrast, citrate is directly metabolized through the TCA cycle, while ferulic acid enters the TCA cycle by succinyl-CoA and acetyl-CoA following the protocatechuate pathway. Ferulic acid is a precursor of the flavor compound vanillin, and many microorganisms have been investigated to engineer vanillin production from ferulic acid (Lesage-Meessen *et al.*, 1996; Muheim and Lerch, 1999; Okeke and Venturi, 1999; Achterholt *et al.*, 2000; Overhage *et al.*, 2003; Peng *et al.*, 2003; Plaggenborg *et al.*, 2006; Barghini *et al.*, 2007; Yoon *et al.*, 2007; Hua *et al.*, 2007; Tilay *et al.*, 2010; Di Gioia *et al.*, 2011; Fleige *et al.*, 2013). *P. putida* KT2440 can efficiently degrade ferulic acid and convert it to vanillin, increasing the potential of this bacteria as production strain (Plaggenborg *et al.*, 2003; Graf and Altenbuchner, 2014). Furthermore, *P. putida* is able to use several amino acids as carbon sources with some compounds being preferred over others. The bacterium preferentially uses proline, alanine, glutamate, glutamine and histidine as energy sources, while valine, isoleucine, leucine, tyrosine, phenylalanine, threonine, glycine and serine are less preferred (Moreno *et al.*, 2009). The general catabolism of amino acids, involves aminotransferases to produce  $\alpha$ -ketoacids, which are afterward decarboxylated (Blatt *et al.*, 1966). Differently from other amino acids, serine is mainly dehydrated to pyruvate by the enzyme L-serine dehydratase, as the typical transamination event produces hydroxypyruvate, a very reactive and toxic molecule (Schneider and Stephens, 1990; Fiedler *et al.*, 2002; Duggleby, 2005; Baykal *et al.*, 2006; Karsten and Cook, 2009). The L-serine dehydratase enzyme

prevents the formation of hydroxypyruvate, but it does not completely prevent serine toxicity as the enzyme can generate the reactive  $\alpha$ -aminoacrylic acid. Other enzymes (e.g. O-acetylserine and P-serine) that use serine or its variants can generate aminoacrylate (de Lorenzo *et al.*, 2015). The known serine toxicity can explain the long lag phase and the lower growth rate observed already for our wt strain in the experiments described here. The transcriptomic data have been used to integrate and construct a realistic genome-scale metabolic model of *P. putida*. The *in silico* growth of *P. putida* KT2440 on the four substrates confirmed the expression value and the metabolic fluxes of the RNA-seq data.

#### *The porins for citrate, ferulic acid and serine uptake*

The transcriptomic data have highlighted substrate uptake systems in *P. putida* (Fig. 4). The results reported here have revealed three predicted outer-membrane porins, OpdH, BenF-like and OprD, which may be involved in the transport of citrate, ferulic acid and serine respectively. They all belong to the OprD family of porins, whose members form canals for several substrates, such as amino acids, carboxylic acid and carbapenems in *P. aeruginosa* (Trias and Nikaido, 1990; Tamber *et al.*, 2006; Li *et al.*, 2012; Skurnik *et al.*, 2013; Shen *et al.*, 2015). The transposon mutant growth results confirmed the involvement of OpdH and BenF-like porins in growth on citrate and ferulic acid, respectively, while the OprD porin was not addressed here. Therefore more studies are required to confirm the association of this porin in serine transport.

OpdH has been previously shown to be involved in tricarboxylate (*cis*-aconitate, isocitrate and citrate) transport in *P. aeruginosa* (Tamber *et al.*, 2006, 2007), and the results obtained here strongly support the possibility of a similar role in *P. putida*. On the other hand, there may be differences between the two *Pseudomonas* species. In the case of OprD,

it has been shown that the porin plays a role in the transport of basic amino acids and carbapenems, such as imipenem, in *P. aeruginosa* (Trias and Nikaido, 1990; Huang and Hancock, 1996; Ochs *et al.*, 1999, 2000; Pirnay *et al.*, 2002; Chevalier *et al.*, 2007; Li *et al.*, 2012; Shen *et al.*, 2015), but no studies have revealed its specific function in serine transport, as this work suggests in *P. putida*. Moreover, a study on *P. putida* (Bojanovic *et al.* manuscript in preparation) did not detect a high expression of *oprD* during imipenem stress, strengthening the idea of a different role of this porin in *P. putida* and *P. aeruginosa*. These concepts underline possible similarities and differences between the two species, and suggest the necessity to further investigate OprD in *P. putida*.

#### *Inner membrane transporters*

In this work inner membrane transport systems for the carbon sources have been revealed (Fig. 4). There are several annotated genes in the *P. putida* KT2440 genome that may be involved in citrate transport, including two genes encoding citrate transporters (CitMHS family) (PP\_0147 and PP\_2057), a gene cluster for the tricarboxylate transporter TctABC (PP\_1418-16), a member of the major facilitator superfamily of transporters (PP\_2703), and a putative citrate transporter (PP\_3074). The analysis presented here identifies the citrate transporter (PP\_0147) of the CitMHS family and the tricarboxylate transporter system TctABC as being involved in citrate uptake.

The CitMHS family includes members that transport citrate in a complex with a bivalent metal ion. The best characterized members are CitM and CitH from *B. subtilis* (Boorsma *et al.*, 1996). CitM binds citrate in complex with  $Mg^{2+}$  and it is the transporter responsible for growth on citrate and isocitrate. CitH transports citrate in complex with

Ca<sup>2+</sup> but it is not induced under the citrate growth condition and its function is still unclear (Krom *et al.*, 2000; Li and Pajor, 2002; Warner and Lolkema, 2002). Other functionally characterized systems of the CitMHS family have been described for *Streptococcus mutans* (Korithoski *et al.*, 2005), *Enterococcus faecalis* (Blancato *et al.*, 2006), *Streptomyces coelicolor* (Lensbouwer *et al.*, 2008), and *Corynebacterium glutamicum* (Brocker *et al.*, 2009).

The *P. putida* TctABC is likely related to the well characterized TctABC system (TTT family) of *S. typhimurium*, which binds citrate by the TctC subunit through a Na<sup>+</sup>-dependent mechanism (Sweet *et al.*, 1979). The system is inducible by citrate or isocitrate, or to a lesser extent, if *cis*-aconitate is present in the medium (Winnen *et al.*, 2003). Tct systems have also been functionally characterized in *Bordetella pertussis* (Antoine *et al.*, 2003) and *Corynebacterium glutamicum* (Brocker *et al.*, 2009), and showed the same function of citrate uptake.

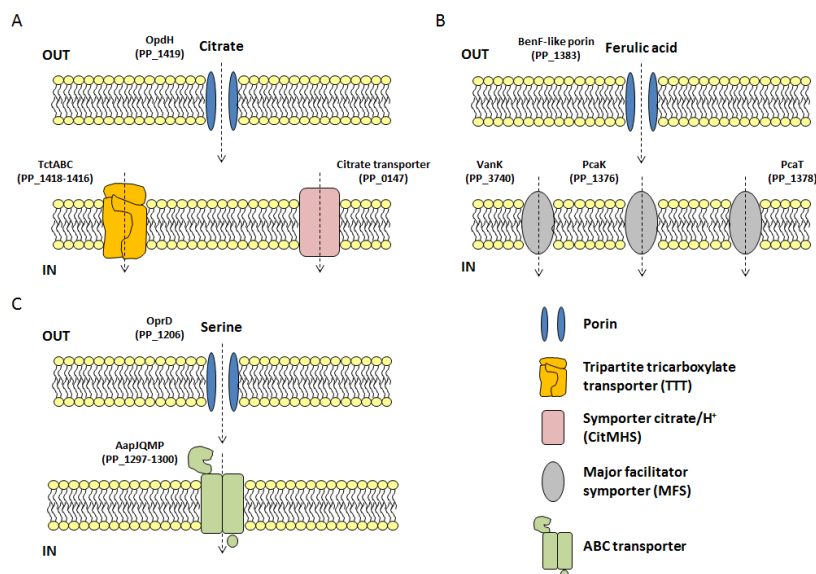
The growth assay with transposon mutants showed that the subunit TctA of the TctABC transporter is involved in *P. putida* cells growing on citrate, as well as the citrate transporter (PP\_0147), leading to two major conclusions. First, TctA is likely the important subunit for citrate transport in the TctABC system. As has been observed previously in *Salmonella*, TctA may be the actual transporter while TctB and TctC confer high affinity to the substrate (Winnen *et al.*, 2003). In this perspective, it would be interesting to test the growth on citrate for strains with disrupted PP\_1417 and PP\_1418 genes, encoding the TctB subunit and the TctC subunit, respectively, which was not done in this study. Second, the citrate transporter may be more active in citrate transport than the TctABC system, as PP\_0147 is more highly expressed and a disrupted gene yielded a longer lag phase prior to exponential growth (Fig. 3A).



For ferulic acid, the transcriptomic data revealed PcaK, PcaT and VanK as possible systems involved in ferulic acid transport. *P. putida* PcaK catalyzes the accumulation of *p*-hydroxybenzoate and protocatechuate, while benzoate is not a PcaK-substrate (Harwood *et al.*, 1994; Nichols and Harwood, 1997). Similar to PcaK, VanK also has specificity for *p*-hydroxybenzoate and protocatechuate in *Acinetobacter* sp. ADP1. However, the complete set of compounds transported by PcaK and VanK remains to be elucidated (D'Argenio *et al.*, 1999). No functional characterization has been performed for PcaT, which remains a candidate  $\beta$ -ketoadipate transporter (Harwood and Parales, 1996; Jiménez *et al.*, 2002). The results of the transposon mutant growth experiments implicate the PcaT and VanK transporters in growth on ferulic acid (PcaK was not tested in this study). Based on their strong involvement in the transport of aromatic substrates and ferulic acid uptake in particular, PcaK, PcaT and VanK are interesting candidates to test in further characterization study and optimization of *P. putida* for vanillate production.

Finally, for serine transporters, the *P. putida* KT2440 genome annotates the serine/threonine transporter SstT (PP\_2443), the D-alanine/D-serine/glycine permease CycA (PP\_4840), and the aromatic amino acid ABC transporter permease SdaC (PP\_3589). The data showed that SstT was not the major player in serine uptake, but rather a high expression of the general polar amino acid ABC transporter AapJQMP was observed instead. Insertions in genes encoding three subunits of the transporter (AapJ, AapQ, AapM), severely affected the growth of *P. putida* on serine. A similar involvement of an ABC transporter in amino acid uptake in *P. putida* was observed previously for growth in glutamate. Indeed, a previous study has characterized the

ABC transporter AatJMQP (PP\_1071-1068) as a selective uptake system for the amino acids glutamate and aspartate (Singh and Röhm, 2008).



**Figure 4.** The uptake systems identified in this study. Representation of porins and inner membrane uptake systems identified in the transport of the carbon sources; citrate (A), ferulic acid (B), and serine (C).

Overall, this work contributes to a better understanding of *P. putida* metabolic versatility and adaptation to different environments. This work included the identification and preliminary characterizations of uptake systems involved in citrate, ferulic acid and serine uptake. The information acquired in this work will serve as the benchmark for further detailed studies to address the specificity and the role of these transporters. The results reported here are required for the global understanding of *P. putida* metabolism and *in silico* modeling

reconstruction, and provide essential knowledge to develop the bacterium as an efficient cell factory.

## **Experimental Procedures**

### *Bacterial strains and growth conditions*

*Pseudomonas putida* strain KT2440 was cultivated in M9 minimal medium (per liter:  $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ , 70 g;  $\text{KH}_2\text{PO}_4$ , 30 g;  $\text{NH}_4\text{Cl}$ , 10 g;  $\text{NaCl}$ , 5 g) supplemented with ammonium iron citrate, magnesium sulfate, and trace metals (per liter:  $\text{H}_3\text{BO}_3$ , 300 mg;  $\text{ZnCl}_2$ , 50 mg;  $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ , 30 mg;  $\text{CoCl}_2$ , 200 mg;  $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$ , 10 mg;  $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$ , 20 mg; and  $\text{NaMoO}_4 \cdot 2\text{H}_2\text{O}$ , 30 mg) (Abril et al., 1989). Bacteria were grown in one of the chosen carbon sources: glucose (0.5% (w/v)), sodium citrate (10 mM), ferulic acid (10 mM), or serine (10 mM). Single colonies isolated from LB agar plates containing 25  $\mu\text{g/mL}$  chloramphenicol and grown overnight at 30°C, were inoculated in 5 mL M9 medium. The overnight cultures were diluted to a starting  $\text{OD}_{600}$  of 0.01 (for ferulic acid and serine) and 0.1 (for glucose and sodium citrate) in 100 mL M9 medium in 250 mL Erlenmeyer flasks. All liquid cultures were grown at 30°C with vigorous shaking at 250 rpm.

### *Cell harvest and RNA isolation*

*P. putida* KT2440 cells were harvested in mid-exponential phase, correspond to  $\text{OD}_{600} \sim 0.6$  for glucose,  $\text{OD}_{600} \sim 0.4$  for citrate,  $\text{OD}_{600} \sim 0.3\text{-}0.4$  for ferulic acid, and  $\text{OD}_{600} \sim 0.15\text{-}0.2$  for serine. Cells harvesting was done by transferring 20 mL of each culture into 50-mL Falcon tubes containing 4 mL of stop solution (5% phenol in 95% ethanol, 4°C), vortexed for 15 sec and kept on ice for 5 min. After centrifugation (3500 g, 10 min, 4°C) in a Multifuge X3 Fr centrifuge (Thermo Scientific), the supernatant was removed and the pellet was frozen in liquid nitrogen and

stored at -80°C. Total RNA extraction and DNase I treatment for DNA removal were performed as described elsewhere (Gomez-Lozano et al., 2012). The integrity of total RNA, and DNA contamination were assessed with a RNA 6000 Nano chip on Agilent 2100 Bioanalyzer (Agilent Technologies). Total RNA was extracted from three biological replicate cultures for each condition.

#### *Library preparation and RNA sequencing*

Sequencing libraries were constructed using the Illumina® TruSeq® Stranded mRNA Sample Preparation kit (Sultan et al., 2012). Libraries were validated on the Agilent 2100 Bioanalyzer using a DNA 1000 chip, and concentration estimated by Qubit 2.0 Fluorometer (Invitrogen, Life Technologies). Libraries were normalized to 10 nM using 10 mM Tris-Cl, pH 8.5, 0.1% Tween 20, and finally pool together (10 µL of each normalized library). The pool was validated with the DNA High Sensitivity Assay on Agilent 2100 Bioanalyzer (Agilent Technologies) and the concentration confirmed on a Qubit 2.0 Fluorometer. The libraries were sequenced twice on the Illumina HiSeq2000 platform (Beckman Coulter Genomics).

#### *Data analysis and differential expression*

The sequencing reads were checked for quality by evaluation of average quality per reads Phred score. For each sample the reads deriving from the two sequencing runs were merged. Then reads were mapped onto the *P. putida* KT2440 genome (RefSeq Accession No. NC\_002947.3) with Bowtie2 by Rockhopper 2.0.3 software (McClure et al., 2013; Tjaden, 2015). RPKM gene counts were used for differential expression analysis with T-REx pipeline (de Jong *et al.*, 2015)

(<http://genome2d.molgenrug.nl>). Data were handled by R Bioconductor and Microsoft Excel.

### *Genome-scale modeling and simulation of growth rates*

All the modeling data and codes have been made available at <https://gitlab.com/joaocardoso/ppu-rna-seq-2016>.

The genome-scale model of *P. putida* KT2440 iJP962 was used (Oberhardt *et al.*, 2011). The model contains 962 genes, 1070 reactions and 897 metabolites. As the iJP962 model lacks the ferulic acid degradation pathway, we included the reactions from another model - iJN746 (J Overhage *et al.*, 1999; Nogales *et al.*, 2008). To complete the pathway, 12 reactions and 11 genes were included in the model (Table S7).

Corrections to the model were done by Script 1: <http://goo.gl/BNvVxw>.

Flux Balance Analysis (FBA) was used to simulate metabolic functions. FBA is an approach that uses linear programming to predict metabolic fluxes. The method assumes a pseudo steady-state where all internal metabolite concentrations remain unchanged over time. The mathematical problem is described as follows:

$$\max: \quad \mathbf{c}^T \cdot \mathbf{v} \quad (1)$$

$$\text{st:} \quad \mathbf{S} \cdot \mathbf{v} = 0 \quad (2)$$

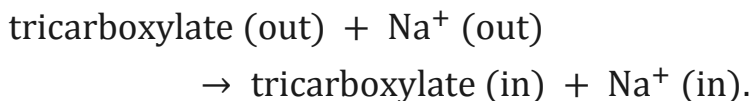
$$\mathbf{v}_{lb} \leq \mathbf{v} \leq \mathbf{v}_{ub} \quad , \forall v \text{ in } M \quad (3)$$

where  $\mathbf{v}$  is the vector of fluxes of the  $M$  reactions in the model. Equation 2 ensures the steady-state assumption, where  $\mathbf{S}$  is the stoichiometric matrix, whose columns and rows correspond to reactions and metabolites respectively; by convention, negative and positive

stoichiometric coefficients are used for substrates and products. Equation 3 defines the upper and lower bounds of each flux (e.g. for irreversible reactions  $v_{lb} = 0$ ). By setting bounds on the model's exchange reactions we can define which metabolites can be taken up and simulate different environmental conditions, e.g., growth in different carbon sources (Orth *et al.*, 2010). We can define the objective (Equation 1) as the biomass reaction – a reaction describing composition of the cells – and then calculate the growth rate.

#### *Tricarboxylate transport in the model*

The general form of the Tricarboxylate transport reactions is:



With this information, we identified eight tricarboxylate molecules present in the model from the ChEBI database (de Matos *et al.*, 2009). For each of those molecules we created a specific reaction. The reaction is associated with the genes PP\_1416, PP\_1417, PP1\_418 and PP\_1419.

#### *Integration of the RNA-Seq data with the model*

The Gene Inactivity Moderated by Metabolism and Expression (GIMME) (Becker and Palsson, 2008) method was used to integrate the RNA-Seq data with the metabolic model. GIMME is a two-step method: the first step requires finding the fluxes through some predefined reactions using FBA; the second step is to minimize the usage of fluxes whose corresponding enzymes are expressed below a given threshold,

while satisfying a set percentage of the original objective value. The second problem is described as:

$$\begin{aligned}
& \min: && c_i |v_i| \\
& \text{st:} && S \cdot v = 0 \\
& && v_{lb} \leq v \leq v_{ub} \quad , \forall v \text{ in } M \\
& \text{where:} && c_i = \begin{cases} x_{cutoff} - x_i & \text{if } x_{cutoff} > x_i \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

The objective is to minimize the number of active fluxes that are not consistent with the transcription data. To do that we penalize the fluxes of reactions carried by enzymes that are not expressed. The penalty is represented by  $c_i$  and it is defined by how much the expression  $x$  is bellow a given threshold  $x_{cutoff}$ . By minimizing the number of penalized reactions, we can calculate the flux distribution that best satisfies the expression data. The expression value  $x$  for each reaction is explained in the (File S1 - *Mapping expression values to reactions*).

Following the optimization, the inconsistencies between the predicted fluxes and the data were calculated by calculating the Inconsistency Score. The Inconsistency Score of a reaction is given by:

$$score = v \cdot (x_{cutoff} - x)$$

A non- zero values means that the activity of the reaction is inconsistent with the transcription data (Becker and Palsson, 2008). After identifying the inconsistent reactions, we apply constraints to the model to include only reactions that are: a) carried by expressed enzymes; b) required for the model to carry flux through the pathways required; and c) include reactions that have no expression information. The implementation of GIMME used in this work can be found in the *driven*

python package (ZENODO, <http://driven.bio>). We did a sensitivity analysis of  $x_{cutoff}$  and used the value 20 (File S1 - *GIMME parameters and inconsistency scores*).

#### *Evaluating active pathways*

After building condition specific models with expression data we used Flux Variability Analysis (FVA) to identify the reactions that can carry flux. FVA consists of maximizing and minimizing the flux through reactions to determine the minimum and maximum possible values under the constraints set for the model (Mahadevan and Schilling, 2003). By enforcing the measured growth rates, we identified which pathways can be used to sustain growth and satisfy the constraints imposed by expression.

#### *Pathway visualization*

We drew the metabolic map using Escher, a web-based tool to draw and visualize biological pathways (King *et al.*, 2015). Escher allows the integration of data into the pathways. We overlaid the flux data predicted using the model and set the color scale to display active pathways in blue and the non-active pathways in grey for each growth condition.

#### *Transposon mutant growths assay*

Mutant mini-Tn5 strains of *P. putida* KT2440 were obtained from the available collection at the Pseudomonas Reference Culture Collection (PRCC) (Duque *et al.*, 2007). The transposon mutants used in this work are listed in Supplement Table S6. Mutant colonies from LB agar plates supplemented with kanamycin (50 µg/ml) were inoculated in 5 mL LB medium and incubated at 30°C with vigorous shaking overnight at 250 rpm. Cells from overnight cultures were harvested (8000 rpm, 4 min,



room temperature) in a Multifuge X3 Fr centrifuge (Thermo Scientific). The pelleted cells were resuspended in 5 mL M9 media followed by OD<sub>600</sub> measurement. Cells were diluted 1:200 in fresh M9 medium with one of the specific carbon sources and plated in a 96-well plate. Cell growth was monitored by a kinetic assay with the microtiter reader (ELx808 BioTek) by monitoring growth via OD<sub>630</sub> measurement every 10 min.

### **Acknowledgments**

The authors thank Juan Luis Ramos for providing the transposon mutant strains. This work was supported by the Novo Nordisk Foundation Center for Biosustainability, and a PhD grant from the People Programme (Marie Curie Actions) of the European Union Seventh Framework Programme FP7-People-2012-ITN, under grant agreement No. 317058, “BACTORY”.

### **Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

### **References**

- Achterholt, S., Priefert, H., and Steinbüchel, A. (2000) Identification of *Amycolatopsis* sp. strain HR167 genes, involved in the bioconversion of ferulic acid to vanillin. *Appl. Microbiol. Biotechnol.* **54**: 799–807.
- Antoine, R., Jacob-Dubuisson, F., Drobecq, H., Willery, E., Lesjean, S., and Loch, C. (2003) Overrepresentation of a Gene Family Encoding Extracytoplasmic Solute Receptors in *Bordetella*. *J. Bacteriol.* **185**: 1470–1474.
- Ashton, D.M., Sweet, G.D., Somers, J.M., and Kay, W.W. (1980) Citrate transport in *Salmonella typhimurium*: studies with 2-fluoro-L-erythro-citrate as a substrate. *Can. J. Biochem.* **58**: 797–803.

Barghini, P., Di Gioia, D., Fava, F., and Ruzzi, M. (2007) Vanillin production using metabolically engineered *Escherichia coli* under non-growing conditions. *Microb. Cell Fact.* **6**: 13.

Baykal, A., Chakraborty, S., Dodoo, A., and Jordan, F. (2006) Synthesis with good enantiomeric excess of both enantiomers of  $\alpha$ -ketols and acetolactates by two thiamin diphosphate-dependent decarboxylases. *Bioorg. Chem.* **34**: 380–393.

Becker, S.A. and Palsson, B.O. (2008) Context-Specific Metabolic Networks Are Consistent with Experiments. *PLoS Comput. Biol.* **4**:

Belda, E., Heck, R.G.A. Van, Lopez-sanchez, M.J., and Cruveiller, S. (2016) The revisited genome of *Pseudomonas putida* KT2440 enlightens its value as a robust metabolic chassis. *Environ. Microbiol.*

Blancato, V.S., Magni, C., and Lolkema, J.S. (2006) Functional characterization and  $\text{Me}^{2+}$  ion specificity of a  $\text{Ca}^{2+}$ -citrate transporter from *Enterococcus faecalis*. *FEBS J.* **273**: 5121–5130.

Blatt, L., Dorer, F.E., and Sallach, H.J. (1966) Occurrence of Hydroxypyruvate-L-Glutamate Transaminase in *Escherichia coli* and Its Separation from Hydroxypyruvate-Phosphate-L-Glutamate Transaminase. *J. Bacteriol.* **92**: 668–675.

Boorsma, A., van der Rest, M.E., Lolkema, J.S., and Konings, W.N. (1996) Secondary Transporters for Citrate and the  $\text{Mg}^{2+}$ -Citrate Complex in *Bacillus subtilis* Are Homologous Proteins. *J. Bacteriol.* **178**: 6216–6222.

Brocker, M., Schaffer, S., Mack, C., and Bott, M. (2009) Citrate Utilization by *Corynebacterium glutamicum* Is Controlled by the CitAB Two-Component System through Positive Regulation of the Citrate Transport Genes *citH* and *tctCBA*. *J. Bacteriol.* **191**: 3869–3880.

Calisti, C., Ficca, A.G., Barghini, P., and Ruzzi, M. (2008) Regulation of ferulic catabolic genes in *Pseudomonas fluorescens* BF13: Involvement of a MarR family regulator. *Appl. Microbiol. Biotechnol.* **80**: 475–483.

del Castillo, T., Duque, E., and Ramos, J.L. (2008) A set of Activators and Repressors Control Peripheral Glucose Pathways in *Pseudomonas putida* To Yield a Common Central Intermediate. *J. Bacteriol.* **190**: 2331–2339.

del Castillo, T., Ramos, J.L., Rodríguez-Herva, J.J., Fuhrer, T., Sauer, U., and Duque, E. (2007) Convergent Peripheral Pathways Catalyze Initial Glucose Catabolism in *Pseudomonas putida*: Genomic and Flux Analysis. *J. Bacteriol.* **189**: 5142–5152.

Chavarría, M., Nickel, P.I., Pérez-Pantoja, D., and De Lorenzo, V. (2013) The Entner-Doudoroff pathway empowers *Pseudomonas putida* KT2440 with a high tolerance to oxidative stress. *Environ. Microbiol.* **15**: 1772–1785.

Chevalier, S., Bodilis, J., Jaouen, T., Barray, S., Feuilloley, M.G.J., and Orange, N. (2007) Sequence diversity of the OprD protein of environmental *Pseudomonas* strains. *Environ. Microbiol.* **9**: 824–835.

Cowles, C.E., Nichols, N.N., and Harwood, C.S. (2000) BenR, a XylS Homologue, Regulates Three Different Pathways of Aromatic Acid Degradation in *Pseudomonas putida*. *J. Bacteriol.* **182**: 6339–6346.

Creecy, J.P. and Conway, T. (2015) Quantitative bacterial transcriptomics with RNA-seq. *Curr. Opin. Microbiol.* **23**: 133–140.

D'Argenio, D.A., Segura, A., Coco, W.M., Bünz, P. V., and Ornston, L.N. (1999) The Physiological Contribution of *Acinetobacter* PcaK, a Transport System That Acts upon Protocatechuate, Can Be Masked by the Overlapping Specificity of VanK. *J. Bacteriol.* **181**: 3505–3515.

Duggleby, R.G. (2005) Suicide inhibition of acetohydroxyacid synthase by hydroxypyruvate. *J. Enzyme Inhib. Med. Chem.* **20**: 1–4.

Duque, E., Molina-Henares, A.J., Torre, J.D. La, Molina-Henares, M.A., Castillo, T. Del, Lam, J., and Ramos, J.L. (2007) Towards a genome-wide mutant library of *Pseudomonas putida* strain KT2440. *Pseudomonas* **5**: 227–251.

Espinosa-Urgel, M., Kolter, R., and Ramos, J.L. (2002) Root colonization by *Pseudomonas putida*: love at first site. *Microbiology* **148**: 341.

Federal Register (1982) Appendix E, Certified host-vector systems. **47**: 17197.

Fernández, M., Conde, S., De La Torre, J.J., Molina-Santiago, C., Ramos, J.-L.L., Duque, E., et al. (2012) Mechanisms of Resistance to Chloramphenicol in *Pseudomonas putida* KT2440. *Antimicrob. Agents Chemother.* **56**: 1001–1009.

Fiedler, E., Thorell, S., Sandalova, T., Golbik, R., König, S., and Schneider, G. (2002) Snapshot of a key intermediate in enzymatic thiamin catalysis: Crystal structure of the alpha-carbanion of (alpha,beta-dihydroxyethyl)-thiamin diphosphate in the active site of transketolase from *Saccharomyces cerevisiae*. *PNAS* **99**: 591–595.

Fleige, C., Hansen, G., Kroll, J., and Steinbüchel, A. (2013) Investigation of the *Amycolatopsis* sp. Strain ATCC 39116 Vanillin Dehydrogenase and Its Impact on the Biotechnical Production of Vanillin. *Appl. Environ. Microbiol.* **79**: 81–90.

Frank, S., Klockgether, J., Hagendorf, P., Geffers, R., Schöck, U., Pohl, T., et al. (2011) *Pseudomonas putida* KT2440 genome update by cDNA sequencing and microarray transcriptomics. *Environ. Microbiol.* **13**: 1309–1326.

Fuhrer, T., Fischer, E., and Sauer, U. (2005) Experimental Identification and Quantification of Glucose Metabolism in Seven Bacterial Species. *J. Bacteriol.* **187**: 1581–1590.

Garrett, E.S., Perlegas, D., and Wozniak, D.J. (1999) Negative Control of Flagellum Synthesis in *Pseudomonas aeruginosa* Is Modulated by the Alternative Sigma Factor AlgT (AlgU). *J. Bacteriol.* **181**: 7401–7404.

Gasson, M.J., Kitamura, Y., McLauchlan, W.R., Narbad, A., Parr, A.J., Parsons, E.L.H., et al. (1998) Metabolism of Ferulic Acid to Vanillin. *J. Biol. Chem.* **273**: 4163–4170.

Di Gioia, D., Luziatelli, F., Negroni, A., Ficca, A.G., Fava, F., and Ruzzi, M. (2011) Metabolic engineering of *Pseudomonas fluorescens* for the production of vanillin from ferulic acid. *J. Biotechnol.* **156**: 309–316.

Graf, N. and Altenbuchner, J. (2014) Genetic engineering of *Pseudomonas putida* KT2440 for rapid and high-yield production of vanillin from ferulic acid. *Appl. Microbiol. Biotechnol.* **98**: 137–149.

Harwood, C.S., Nichols, N.N., Kim, M.K., Ditty, J.L., and Parales, R.E. (1994) Identification of the pcaRKF Gene Cluster from *Pseudomonas putida*: Involvement in Chemotaxis, Biodegradation, and Transport of 4-Hydroxybenzoate. *J. Bacteriol.* **176**: 6479–6488.

- Harwood, C.S. and Parales, R.E. (1996) The B -Ketoadipate Pathway and the Biology of Self-Identity. *Annu. Rev. Microbiol.* **50**: 553–590.
- Hua, D., Ma, C., Song, L., Lin, S., Zhang, Z., Deng, Z., and Xu, P. (2007) Enhanced vanillin production from ferulic acid using adsorbent resin. *Appl. Microbiol. Biotechnol.* **74**: 783–790.
- Huang, H. and Hancock, R.E.W. (1996) The role of Specific Surface Loop Regions in Determining the Function of the Imipenem-Specific Pore Protein OprD of *Pseudomonas aeruginosa*. *J. Bacteriol.* **178**: 3085–3090.
- Jiménez, J.I., Miñambres, B., García, J.L., and Díaz, E. (2002) Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**: 824–841.
- Johnson, D.A., Tetu, S.G., Phillippy, K., Chen, J., Ren, Q., and Paulsen, I.T. (2008) High-throughput phenotypic characterization of *Pseudomonas aeruginosa* membrane transport genes. *PLoS Genet.* **4**:
- de Jong, A., van der Meulen, S., Kuipers, O.P., and Kok, J. (2015) T-REx: Transcriptome analysis webserver for RNA-seq Expression data. *BMC Genomics* **16**: 663.
- Karsten, W.E. and Cook, P.F. (2009) Detection of a gem-diamine and a stable quinonoid intermediate in the reaction catalyzed by serine-glyoxylate aminotransferase from *Hyphomicrobium methylovorum*. *BBA - Gen. Subj.* **1790**: 575–580.
- Keith, L.M.W. and Bender, C.L. (1999) AlgT (sigma22) Controls Alginate Production and Tolerance to Environmental Stress in *Pseudomonas syringae*. *J. Bacteriol.* **181**: 7176–7184.
- Kim, J., Oliveros, J.C., Nikel, P.I., de Lorenzo, V., and Silva-Rocha, R. (2013) Transcriptomic fingerprinting of *Pseudomonas putida* under alternative physiological regimes. *Environ. Microbiol. Rep.* **5**: 883–891.
- King, Z.A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N.E., and Palsson, B.O. (2015) Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *PLoS Comput. Biol.* **11**: 1–13.

- Korithoski, B., Krastel, K., and Cvitkovitch, D.G. (2005) Transport and Metabolism of Citrate by *Streptococcus mutans*. *J. Bacteriol.* **187**: 4451–4456.
- Krom, B.P., Warner, J.B., Konings, W.N., and Lolkema, J.S. (2000) Complementary Metal Ion Specificity of the Metal-Citrate Transporters CitM and CitH of *Bacillus subtilis*. *J. Bacteriol.* **182**: 6374–6381.
- Lensbouer, J.J., Patel, A., Sirianni, J.P., and Doyle, R.P. (2008) Functional Characterization and Metal Ion Specificity of the Metal-Citrate Complex Transporter from *Streptomyces coelicolor*. *J. Bacteriol.* **190**: 5616–5623.
- Lesage-Meessen, L., Delattre, M., Haon, M., Thibault, J.F., Ceccaldi, B.C., Brunerie, P., and Asther, M. (1996) A two-step bioconversion process for vanillin production from ferulic acid combining *Aspergillus niger* and *Pycnoporus cinnabarinus*. *J. Biotechnol.* **50**: 107–113.
- Li, H., Luo, Y.F., Williams, B.J., Blackwell, T.S., and Xie, C.M. (2012) Structure and function of OprD protein in *Pseudomonas aeruginosa*: From antibiotic resistance to novel therapies. *Int. J. Med. Microbiol.* **302**: 63–68.
- Li, H. and Pajor, A.M. (2002) Functional characterization of CitM, the Mg<sup>2+</sup>-citrate transporter. *J. Membr. Biol.* **185**: 9–16.
- Loeschcke, A. and Thies, S. (2015) *Pseudomonas putida*—a versatile host for the production of natural products. *Appl. Microbiol. Biotechnol.* **99**: 6197–6214.
- de Lorenzo, V., Sekowska, A., and Danchin, A. (2015) Chemical reactivity drives spatiotemporal organisation of bacterial metabolism. *FEMS Microbiol. Rev.* **39**: 96–119.
- Lugtenberg, B.J.J., Dekkers, L., and Bloemberg, G. V (2001) Molecular Determinants of Rhizosphere Colonization by *Pseudomonas*. *Ann. Rev. Phytopathol.* **39**: 461–490.
- Mahadevan, R. and Schilling, C.H. (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* **5**: 264–276.
- Martins dos Santos, V.A.P., Heim, S., Moore, E.R.B., Strätz, M., and Timmis, K.N. (2004) Insights into the genomic basis of niche specificity of *Pseudomonas putida* KT2440. *Environ. Microbiol.* **6**: 1264–1286.

de Matos, P., Alcántara, R., Dekker, A., Ennis, M., Hastings, J., Haug, K., et al. (2009) Chemical Entities of Biological Interest: an update. *Nucleic Acids Res.* **38**: 249–254.

McCloskey, D., Palsson, B.Ø., and Feist, A.M. (2013) Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Mol. Syst. Biol.* **9**: 661.

McClure, R., Balasubramanian, D., Sun, Y., Bobrovskyy, M., Sumby, P., Genco, C.A., et al. (2013) Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res.* **41**: e140.

Moreno, R., Martínez-Gomariz, M., Yuste, L., Gil, C., and Rojo, F. (2009) The *Pseudomonas putida* Crc global regulator controls the hierarchical assimilation of amino acids in a complete medium: Evidence from proteomic and genomic analyses. *Proteomics* **9**: 2910–2928.

Muheim, A. and Lerch, K. (1999) Towards a high-yield bioconversion of ferulic acid to vanillin. *Appl. Microbiol. Biotechnol.* **51**: 456–461.

Nakazawa, T. (2002) Travels of a *Pseudomonas*, from Japan around the world. *Environ. Microbiol.* **4**: 782–786.

Narbad, A. and Gasson, M.J. (1998) Metabolism of ferulic acid via vanillin using a novel CoA-dependent pathway in a newly-isolated strain of *Pseudomonas fluorescens*. *Microbiology* **144**: 1397–1405.

Nelson, K.E., Weinell, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Martins dos Santos, V.A.P., et al. (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**: 799–808.

Newman, E.B. and Walker, C. (1982) L-Serine Degradation in *Escherichia coli* K-12: a Combination of L-Serine, Glycine, and Leucine Used as a Source of Carbon. *J. Bacteriol.* **151**: 777–782.

Nichols, N.N. and Harwood, C.S. (1997) PcaK, a High-Affinity Permease for the Aromatic Compounds 4-Hydroxybenzoate and Protocatechuate from *Pseudomonas putida*. *J. Bacteriol.* **179**: 5056–5061.

Nikel, P.I., Chavarría, M., Fuhrer, T., Sauer, U., and de Lorenzo, V. (2015) *Pseudomonas putida* KT2440 Strain Metabolizes Glucose through a Cycle

Formed by Enzymes of the Entner-Doudoroff, Embden-Meyerhof-Parnas, and Pentose Phosphate Pathways. *J. Biol. Chem.* **290**: 25920–25932.

Nikel, P.I., Kim, J., and de Lorenzo, V. (2013) Metabolic and regulatory rearrangements underlying glycerol metabolism in *Pseudomonas putida* KT2440. *Environ. Microbiol.* **16**: 239–254.

Nishikawa, Y., Yasumi, Y., Noguchi, S., Sakamoto, H., and Nikawa, J. (2008) Functional Analyses of *Pseudomonas putida* Benzoate Transporters Expressed in the Yeast *Saccharomyces cerevisiae*. *Biosci. Biotechnol. Biochem.* **72**: 2034–2038.

Nogales, J., Palsson, B.Ø., and Thiele, I. (2008) A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iJN746 as a cell factory. *BMC Syst. Biol.* **2**: 79.

Oberhardt, M.A., Puchalka, J., dos Santos, V.A.P.M., and Papin, J.A. (2011) Reconciliation of Genome-Scale Metabolic Reconstructions for Comparative Systems Analysis. *PLoS Comput. Biol.* **7**:

Ochs, M.M., Bains, M., and Hancock, R.E.W. (2000) Role of Putative Loops 2 and 3 in Imipenem Passage through the Specific Porin OprD of *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* **44**: 1983–1985.

Ochs, M.M., Lu, C., Hancock, R.E.W., and Abdelal, A.T. (1999) Amino Acid-Mediated Induction of the Basic Amino Acid-Specific Outer Membrane Porin OprD from *Pseudomonas aeruginosa*. *J. Bacteriol.* **181**: 5426–5432.

Okeke, B.C. and Venturi, V. (1999) Construction of Recombinants *Pseudomonas putida* BO14 and *Escherichia coli* QEFCA8 for Ferulic Acid Biotransformation to Vanillin. *J. Biosci. Bioeng.* **88**: 103–106.

Orth, J.D., Thiele, I., and Palsson, B.O. (2010) What is flux balance analysis? *Nat. Biotechnol.* **28**: 245–248.

Overhage, J., Priefert, H., and Steinbüchel, A. (1999) Biochemical and Genetic Analyses of Ferulic Acid Catabolism in *Pseudomonas* sp. Strain HR199. *Appl. Environ. Microbiol.* **65**: 4837–4847.

Overhage, J., Steinbüchel, A., and Priefert, H. (2003) Highly Efficient Biotransformation of Eugenol to Ferulic Acid and Further Conversion to



Vanillin in Recombinant Strains of *Escherichia coli*. *Appl. Environ. Microbiol.* **69**: 6569–6576.

Palleroni, N.J. and Moore, E.R.B. (1986) Taxonomy of the Pseudomonads: experimental approaches. *Bact.* **1**: 3–20.

Parke, D., D'Argenio, D.A., and Ornston, L.N. (2000) Bacteria Are Not What They Eat: That Is Why They Are So Diverse. *J. Bacteriol.* **182**: 257–263.

Parke, D., Garcia, M. a, and Ornston, L.N. (2001) Cloning and Genetic Characterization of *dca* Genes Required for  $\beta$  -Oxidation of Straight-Chain Dicarboxylic Acids in *Acinetobacter* sp . Strain ADP1. *Appl. Environ. Microbiol.* **67**: 4817–4827.

Parthasarathy, S., Lu, F., Zhao, X., Li, Z., Bain, K., Rutter, M.E., et al. (2010) Structure of a putative BenF-like porin from *Pseudomonas fluorescens* Pf-5 at 2.6 Å resolution. *Proteins Struct. Funct. Bioinforma.* **78**: 3056–3062.

Peng, X., Misawa, N., and Harayama, S. (2003) Isolation and Characterization of Thermophilic Bacilli Degrading Cinnamic, 4-Coumaric, Ferulic Acids. *Appl. Environ. Microbiol.* **69**: 1417–1427.

Pirnay, J.P., De Vos, D., Mossialos, D., Vanderkelen, A., Cornelis, P., and Zizi, M. (2002) Analysis of the *Pseudomonas aeruginosa oprD* gene from clinical and environmental isolates. *Environ. Microbiol.* **4**: 872–882.

Plaggenborg, R., Overhage, J., Loos, A., Archer, J.A.C., Lessard, P., Sinskey, A.J., et al. (2006) Potential of *Rhodococcus* strains for biotechnological vanillin production from ferulic acid and eugenol. *Appl. Microbiol. Biotechnol.* **72**: 745–755.

Plaggenborg, R., Overhage, J., Steinbüchel, a, and Priefert, H. (2003) Functional analyses of genes involved in the metabolism of ferulic acid in *Pseudomonas putida* KT2440. *Appl. Microbiol. Biotechnol.* **61**: 528–535.

Poblete-Castro, I., Becker, J., Dohnt, K., dos Santos, V.M., and Wittmann, C. (2012) Industrial biotechnology of *Pseudomonas putida* and related species. *Appl. Microbiol. Biotechnol.* **93**: 2279–2290.

Puchalka, J., Oberhardt, M.A., Godinho, M., Bielecka, A., Regenhart, D., Timmis, K.N., et al. (2008) Genome-Scale Reconstruction and Analysis of the

*Pseudomonas putida* KT2440 Metabolic Network Facilitates Applications in Biotechnology. *PLoS Comput. Biol.* **4**.

Ramos-González, M.I. and Molin, S. (1998) Cloning, sequencing, and phenotypic characterization of the *rpoS* gene from *Pseudomonas putida* KT2440. *J. Bacteriol.* **180**: 3421–3431.

Raskó, I. and Alföldi, L. (1971) Biosynthetic L-Threonine Deaminase as the Origin of L-Serine Sensitivity of *Escherichia coli*. *Eur. J. Biochem.* **21**: 424–427.

Regenhardt, D., Heuer, H., Heim, S., Fernandez, D.U., Strömpl, C., Moore, E.R.B., and Timmis, K.N. (2002) Pedigree and taxonomic credentials of *Pseudomonas putida* strain KT2440. *Environ. Microbiol.* **4**: 912–915.

Revelles, O., Wittich, R.-M., and Ramos, J.L. (2007) Identification of the initial steps in d-lysine catabolism in *Pseudomonas putida*. *J. Bacteriol.* **189**: 2787–2792.

Rojo, F. (2010) Carbon catabolite repression in *Pseudomonas*: optimizing metabolic versatility and interactions with the environment. *FEMS Microbiol. Rev.* **34**: 658–684.

La Rosa, R., Behrends, V., Williams, H.D., Bundy, J.G., and Rojo, F. (2015) Influence of the Crc regulator on the hierarchical use of carbon sources from a complete medium in *Pseudomonas*. *Environ. Microbiol.* **18**: 807–18.

La Rosa, R., Nogales, J., and Rojo, F. (2015) The Crc/CrcZ-CrcY global regulatory system helps the integration of gluconeogenic and glycolytic metabolism in *Pseudomonas putida*. *Environ. Microbiol.* **17**: 3362–3378.

Saravolac, E.G., Taylor, N.F., Benz, R., and Hancock, R.E.W. (1991) Purification of Glucose-Inducible Outer Membrane Protein OprB of *Pseudomonas putida* and Reconstitution of Glucose-Specific Pores. *J. Bacteriol.* **173**: 4970–4976.

Schneider, T.D. and Stephens, R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**: 6097–6100.

Schnider-Keel, U., Lejbølle, K.B., Baehler, E., Haas, D., and Keel, C. (2001) The Sigma Factor AlgU (AlgT) Controls Exopolysaccharide Production and

Tolerance towards Desiccation and Osmotic Stress in the Biocontrol Agent *Pseudomonas fluorescens* CHA0. *Appl. Environ. Microbiol.* **67**: 5683–5693.

Shen, J., Pan, Y., and Fang, Y. (2015) Role of the Outer Membrane Protein OprD2 in Carbapenem-Resistance Mechanisms of *Pseudomonas aeruginosa*. *PLoS One* **10**: 1–9.

Singh, B. and Röhm, K.-H. (2008) Characterization of a *Pseudomonas putida* ABC transporter (AatJMQP) required for acidic amino acid uptake: biochemical properties and regulation by the Aau two-component system. *Microbiology* **154**: 797–809.

Skurnik, D., Roux, D., Cattoir, V., Danilchanka, O., Lu, X., Yoder-Himes, D.R., et al. (2013) Enhanced in vivo fitness of carbapenem-resistant *oprD* mutants of *Pseudomonas aeruginosa* revealed through high-throughput sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **110**: 20747–20752.

Sohn, S.B., Kim, T.Y., Park, J.M., and Lee, S.Y. (2010) In silico genome-scale metabolic analysis of *Pseudomonas putida* KT2440 for polyhydroxyalkanoate synthesis, degradation of aromatics and anaerobic survival. *Biotechnol. J.* **5**: 739–750.

Somers, J.M., Sweet, G.D., and Kay, W.W. (1981) Fluorocitrate resistant tricarboxylate transport mutants of *Salmonella typhimurium*. *Mol. Gen. Genet.* **181**: 338–45.

Sudarsan, S., Dethlefsen, S., Blank, L.M., Siemann-Herzberg, M., and Schmid, A. (2014) The Functional Structure of Central Carbon Metabolism in *Pseudomonas putida* KT2440. *Appl. Environ. Microbiol.* **80**: 5292–5303.

Sweet, G.D., Kay, C.M., and Kay, W.W. (1984) Tricarboxylate-binding Proteins of *Salmonella typhimurium*. *J. Biol. Chem.* **259**: 1586–1592.

Sweet, G.D., Somers, J.M., and Kay, W.W. (1979) Purification and properties of a citrate-binding transport component, the C protein of *Salmonella typhimurium*. *Can. J Biochem.* **57**: 710–715.

Tamber, S., Maier, E., Benz, R., and Hancock, R.E.W. (2007) Characterization of OpdH, a *Pseudomonas aeruginosa* Porin Involved in the Uptake of Tricarboxylates. *J. Bacteriol.* **189**: 929–939.

Tamber, S., Ochs, M.M., and Hancock, R.E.W. (2006) Role of the Novel OprD Family of Porins in Nutrient Uptake in *Pseudomonas aeruginosa*. *J. Bacteriol.* **188**: 45–54.

Tilay, A., Bule, M., and Annapure, U. (2010) Production of Biovanillin by One-Step Biotransformation Using Fungus *Pycnoporous cinnabarinus*. *J. Agric. Food Chem.* **58**: 4401–4405.

Timmis, K.N. (2002) *Pseudomonas putida*: a cosmopolitan opportunist par excellence. *Environ. Microbiol.* **4**: 779–781.

Tiso, T., Wierckx, N., and Blank, L. (2014) Non-Pathogenic *Pseudomonas* as Platform for Industrial Biocatalysis.

Tjaden, B. (2015) De novo assembly of bacterial transcriptomes from RNA-seq data. *Genome Biol.* **16**: 1.

Trias, J. and Nikaido, H. (1990) Protein D2 Channel of the *Pseudomonas aeruginosa* Outer Membrane Has a Binding Site for Basic Amino Acids and Peptides. *J. Biol. Chem.* **265**: 15680–15684.

Velázquez, F., Pflüger, K., Cases, I., De Eugenio, L.I., and De Lorenzo, V. (2007) The Phosphotransferase System Formed by PtsP, PtsO, and PtsN Proteins Controls Production of Polyhydroxyalkanoates in *Pseudomonas putida*. *J. Bacteriol.* **189**: 4529–4533.

Venturi, V., Zennaro, F., Degrassi, G., Okeke, B.C., and Bruschi, C. V. (1998) Genetics of ferulic acid bioconversion to protocatechuic acid in plant-growth-promoting *Pseudomonas putida* WCS358. *Microbiology* **144**: 965–973.

Vílchez, S., Molina, L., Ramos, C., and Ramos, J.L. (2000) Proline Catabolism by *Pseudomonas putida*: Cloning, Characterization, and Expression of the put Genes in the Presence of Root Exudates. *J. Bacteriol.* **182**: 91.

Walshaw, D.L. and Poole, P.S. (1996) The general L-amino acid permease of *Rhizobium leguminosarum* is an ABC uptake system that also influences efflux of solutes. *Mol. Microbiol.* **21**: 1239–1252.

Wang, Q. and Nomura, C.T. (2010) Monitoring differences in gene expression levels and polyhydroxyalkanoate (PHA) production in *Pseudomonas putida* KT2440 grown on different carbon sources. *J. Biosci. Bioeng.* **110**: 653–659.

Warner, J.B. and Lolkema, J.S. (2002) Growth of *Bacillus subtilis* on citrate and isocitrate is supported by the Mg<sup>2+</sup>-citrate transporter CitM. *Microbiology* **148**: 3405–3412.

Widenhorn, K.A., Somers, J.M., and Kay, W.W. (1988) Expression of the Divergent Tricarboxylate Transport Operon (*tctI*) of *Salmonella typhimurium*. *170*: 3223–3227.

Widenhorn, K.A., Somers, J.M., and Kay, W.W. (1989) Genetic Regulation of the Tricarboxylate Transport Operon (*tctI*) of *Salmonella typhimurium*. *J. Bacteriol.* **171**: 4436–4441.

Winnen, B., Hvorup, R.N., and Saier, M.H. (2003) The tripartite tricarboxylate transporter (TTT) family. *Res. Microbiol.* **154**: 457–465.

Wylie, J.L. and Worobec, E.A. (1995) The OprB Porin Plays a Central Role in Carbohydrate Uptake in *Pseudomonas aeruginosa*. *J. Bacteriol.* **177**: 3021–3026.

Yoon, S.H., Lee, E.G., Das, A., Lee, S.H., Li, C., Ryu, H.K., et al. (2007) Enhanced vanillin production from recombinant *E. coli* using NTG mutagenesis and adsorbent resin. *Biotechnol. Prog.* **23**: 1143–1148.

## Supplementary Information

**Figure S1.** Representation of the four metabolic pathways described by our expression data.

**Figure S2.** Metabolic modeling and fluxes analysis of growth in glucose.

**Figure S3.** Metabolic modeling and fluxes analysis of growth in citrate.

**Figure S4.** Metabolic modeling and fluxes analysis of growth in ferulic acid.

**Figure S5.** Metabolic modeling and fluxes analysis of growth in serine.

**Figure S6.** Gene-protein-reaction association.

**Figure S7.** Sensitivity plots for GIMME parameterization.

Supplementary Figures S1, S2, S3, S4, S5 are available at <https://gitlab.com/joaocardoso/ppu-rna-seq-2016/tree/master/images>.

**Table S1.** Mapping statistics.

**Table S2.** Common upregulated genes in glucose versus citrate, ferulic acid and serine. (Not reported in this thesis as too long).

**Table S3.** Differential expressed genes in citrate versus glucose. (Not reported in this thesis as too long).

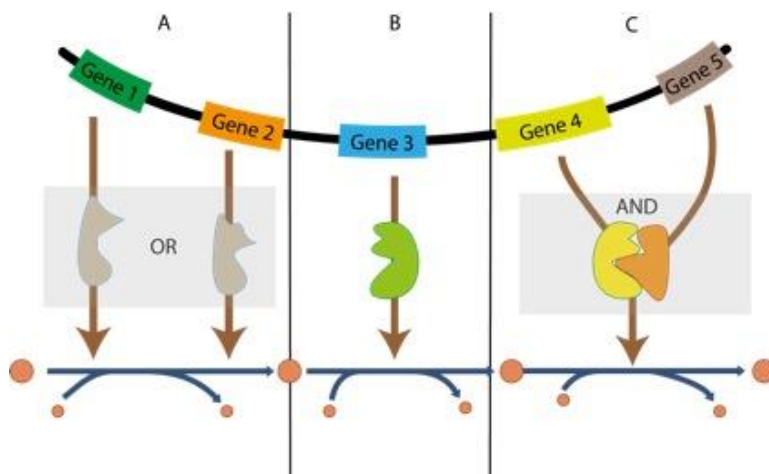
**Table S4.** Differential expressed genes in ferulic acid versus glucose. (Not reported in this thesis as too long).

**Table S5.** Differential expressed genes in serine versus glucose. (Not reported in this thesis as too long).

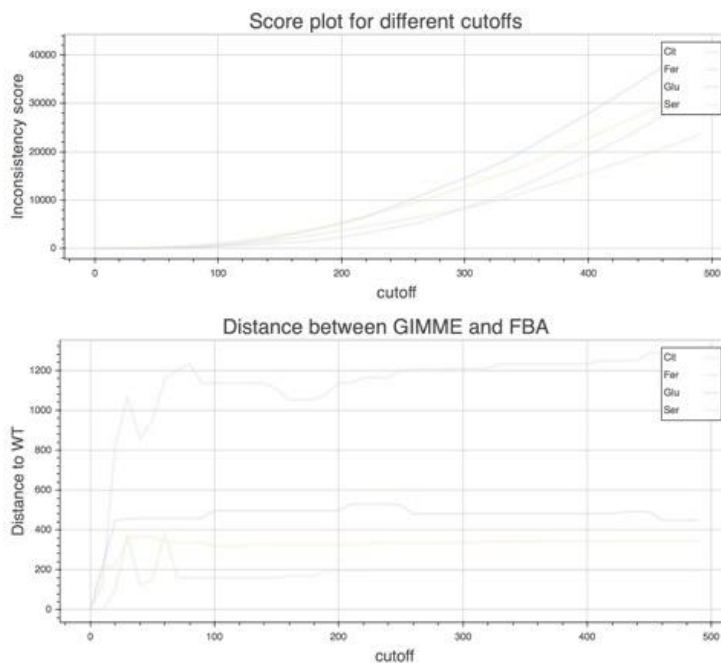
**Table S6.** List of the strains used in this study.

**Table S7.** Genes and reactions included for ferulic acid pathway

**File S1.** Supplementary Methods



**Figure S6.** Gene-protein-reaction association. The relation between genes and reactions in the GEM are described by Boolean relationships. These three examples describe the general cases. (A) Two or more isozymes can carry the same metabolic reactions. When the one of the genes is knocked out, the reaction can still occur. (B) A single gene encodes an enzyme. If the gene is silenced, there is no flux through the reaction. (C) Two or more enzymes are required to carry on a function. For example, enzymatic complexes with multiple subunits cannot work if one of the subunits is not present.



**Figure S7.** Sensitivity plots for GIMME parameterization. We evaluated how changing the threshold affects the results obtained for GIMME. We tested how much it differs from a normal FBA simulation (A) and how the inconsistency scores change (B).



**Table S1.** Mapping statistics.

The total number of reads per sample obtained after merging the reads generated with the two sequencing runs.

The mapping statistics on the *Pseudomonas putida* KT2440 genome are reported: number of reads mapping to rRNA, number of reads mapping to protein-coding genes, and number of reads mapping to unannotated regions.

Sample	Glu1	Glu2	Glu3	Cit1	Cit2	Cit3
Total reads	70657144	21089297	21996209	57982776	58705659	71184603
Mapped reads (%)	64582810 (91%)	18655054 (88%)	19532084 (89%)	50727967 (87%)	58705659 (86%)	64585033 (91%)
rRNA reads (**)	58124529 (90%)	15670245 (84%)	17578876 (90%)	46162450 (91%)	54009206 (92%)	58126530 (90%)
Protein-coding genes (**)	1937484 (3%)	559652 (3%)	585963 (3%)	1521839 (3%)	1761170 (3%)	1937551 (3%)
Intergenic regions (**)	3874969 (6%)	932753 (5%)	585963 (3%)	3043678 (6%)	2935283 (5%)	4520952 (7%)

Sample	Fer1	Fer2	Fer3	Ser1	Ser2	Ser3
Total reads	49695257	46156553	40887634	26101542	63690432	44614433
Mapped reads (%)	43394347 (87%)	40471602 (88%)	36769640 (90%)	23506927 (90%)	58000319 (91%)	34628635 (78%)
rRNA reads (**)	38620969 (89%)	36424442 (90%)	33460372 (91%)	19275680 (82%)	48140265 (83%)	29434340 (85%)
Protein-coding genes (**)	1301830 (3%)	1214148 (3%)	735392 (2%)	940277 (4%)	2320013 (4%)	692573 (2%)
Intergenic regions (**)	3471548 (8%)	2428296 (6%)	2206178 (6%)	3055901 (13%)	6960038 (12%)	4155436 (12%)

\* The % of mapped reads relative to the total number of reads.

\*\* The % of reads relative to the total number of mapped reads.

Abbreviations: Glu, glucose; Cit, citrate; Fer, ferulic acid; Ser, serine.

**Table S6.** List of the strains used in this study.

Strain	Genotype	Ref.
<i>P. putida</i> KT2440	<i>rmo- mod+</i>	DSMZ
Mutant*	Gene interrupted	Ref.
<b>mut::<i>opdH</i></b>	insertion on PP_1419	PRCC
<b>mut::<i>citrate transporter</i></b>	insertion on PP_0147	PRCC
<b>mut::<i>tctA</i></b>	insertion on PP_1416	PRCC
<b>mut::<i>benF</i>-like porin</b>	insertion on PP_1383	PRCC
<b>mut::<i>pcaT</i></b>	insertion on PP_1378	PRCC
<b>mut::<i>vank</i></b>	insertion on PP_3740	PRCC
<b>mut::<i>aapJ</i></b>	insertion on PP_1297	PRCC
<b>mut::<i>aapQ</i></b>	insertion on PP_1298	PRCC
<b>mut::<i>aapM</i></b>	insertion on PP_1299	PRCC

\* Mutants are isogenic to *P. putida* KT2440R strain (M. Espinosa-Urgel *et al.* (2004). Cell Density-Dependent Gene Contributes to Efficient Seed Colonization by *Pseudomonas putida* KT2440. *Appl. Environ. Microbiol.* **70**(9)), a spontaneous rifampicin-resistant derivative of KT2440.

**Table S7.** Genes and reactions included for ferulic acid pathway.

Reaction id	Reaction Name	Genes	Equation
<b>ferulcoas</b>	Feruloyl-CoA synthase	PP_3356	Ferulate+ ATP + CoA → Feruloyl-CoA + AMP + Diphosphate
<b>fcoaha_a</b>	Enoyl-CoA hydratase/aldolase (step 1)	PP_3358	H2O + Feruloyl-CoA → 3- Hydroxy-3-(4-hydroxy-3- methoxyphenyl)propionyl- CoA
<b>fcoaha_b</b>	Enoyl-CoA hydratase/aldolase (step 2)	PP_3358	3-Hydroxy-3-(4-hydroxy-3- methoxyphenyl)propionyl- CoA → Vanillin + Acetyl- CoA
<b>hhmppcoarx</b>	3-Hydroxy-3-(4-hydroxy-3- methoxyphenyl)propionyl-CoA reductase	PP_3354	3-Hydroxy-3-(4-hydroxy-3- methoxyphenyl)propionyl- CoA + NAD → NADH + 4-hydroxy-3- methoxyphenyl-beta- ketopropionyl-CoA
<b>hmpkppncoakt</b>	Beta-Ketothiolase	PP_3355 PP_3754	4-hydroxy-3- methoxyphenyl-beta- ketopropionyl-CoA + CoA → Acetyl-CoA + Vanillyl- CoA
<b>valcoah</b>	Vanillyl-CoA hydratase	PP_3354	Vanillyl-CoA + H2O →

			Vanillate + CoA
<b>vntdm</b>	Vanillate:Oxygen	PP_3737	H+ + O2 + NADH +
	oxidoreductase (demethylating)	PP_3736	Vanillate → Formaldehyde + 3,4-Dihydroxybenzoate + NAD + H2O
<b>vndh</b>	Vanillin:NAD oxidoreductase	PP_3357	Vanillin + NAD + H2O → Vanillate + 2 H+ + NADH
<b>fer_simport_t</b>	Ferulate transport via major facilitator family	PP_3740	Ferulate[out] → Ferulate
<b>fer_mff_t</b>	Benzoate transporter	PP_1376	Ferulate[out] + H+[out] → Ferulate + H+
<b>fer_simp_facilitator_t</b>	Ferulate/H+ symporter via major facilitator superfamily	PP_1378	Ferulate[out] + H+[out] → Ferulate + H+
<b>EX_fer_e</b>	Ferulate exchange		→ Ferulate[out]

## **File S1. Supplementary Methods**

### *Mapping expression values to reactions*

The expression values need to be mapped to the reactions in the model. The relationship between the genotype and the phenotype is described in Fig. S6.

In the first case (Fig. S6B) a single gene is translated into a single protein that catalyze on or more reactions. The expression value for these reactions is the same as the gene. In the second case (Fig. S6A) two, or more, different genes encode different enzymes that have the same functions, also known as isozymes. In this case the expression value for the reactions associated is the maximum of the expression of all genes. The last case (Fig. S6C) describes an enzyme complex where multiple peptides originating from different genes are required in order to the enzyme to function and the expression value assigned to the associated reactions is the minimum expression value (King et al., 2015).

### *GIMME parameters and inconsistency scores*

GIMME requires a cutoff value to define which reactions are active or inactive. This expression value is required because of the noise in the RNA-seq data. To make ensure the quality of the condition specific models we analyzed how different cutoff. We calculated the distance between FBA and GIMME results and the inconsistency scores for a cutoff value between 0 and 500 (Fig. S7).

King, Z. A. *et al.* (2015). Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *PLOS Computational Biology*, **11**(8).

